

## MAIZE PATHOGENESIS-RELATED POLYNUCLEOTIDE AND METHODS OF USE

### CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No.  
5 60/195,801, filed April 10, 2000, which is hereby incorporated in its entirety by  
reference.

### TECHNICAL FIELD

10 The present invention relates generally to plant molecular biology. More  
specifically, it relates to nucleic acids and methods for modulating their expression in  
plants and to transforming genes into plants in order to enhance disease resistance.

### BACKGROUND OF THE INVENTION

15 Disease in plants is produced by biotic and abiotic causes. Biotic causes include  
fungi, viruses, insects, bacteria, and nematodes. Of these, fungi are the most frequent  
causative agents of disease in plants. Abiotic causes of disease in plants include extremes  
of temperature, water, oxygen, soil pH, plus nutrient-element deficiencies and  
imbalances, excess heavy metals, and air pollution.

20 Plant disease outbreaks have resulted in catastrophic crop failures that have  
triggered famines and caused major social change. Generally, the best strategy for plant  
disease control is to use resistant cultivars selected or developed by plant breeders for this  
purpose. Typically, this involves elaborate breeding to incorporate natural resistance  
mechanisms into elite breeding material. The sources of this natural resistance were  
25 often otherwise undesirable plant materials, and so extensive backcrossing and  
RTA01/2093378v1

introgression was needed to recreate the desired background with the disease resistance. Sometimes even this was not obtained, as the resistance mechanism(s) were polygenic. In short, improving disease resistance by conventional breeding is expensive in both time and money. Accordingly, molecular methods are needed to supplement traditional  
5 breeding methods to protect plants from pathogen attack.

A host of cellular processes enables plants to defend themselves from disease caused by pathogenic agents. These processes apparently form an integrated set of resistance mechanisms that is activated by initial infection and then limits further spread of the invading pathogenic microorganism.

10 Subsequent to recognition of a potentially pathogenic microbe, plants can activate an array of biochemical responses. Generally, the plant responds by inducing several local responses in the cells immediately surrounding the infection site. The most common resistance response observed in both nonhost and race-specific interactions is termed the “hypersensitive response” (HR). In the hypersensitive response, cells  
15 contacted by the pathogen, and often neighboring cells, rapidly collapse and dry, producing a necrotic fleck. Other responses include the deposition of callose, the physical thickening of cell walls by lignification, and the synthesis of various antibiotic small molecules and proteins. Genetic factors in both the host and the pathogen determine the specificity of these local responses, which can be very effective in limiting  
20 the spread of infection.

Pathogenesis-related proteins, which have been described in a number of plants (see Bowles (1990) *Ann Rev Biochem* 59:873-907 for review), include the PR-1 proteins. Although their biochemical functions remain unknown, expression of PR-1 proteins is generally induced by pathogens and many abiotic treatments associated with the  
25 elicitation of the defense response, more particularly a hypersensitive response (see WO 89/02437 for a review). These observations make maize PR-1 genes and their promoters ideal candidates for use in the development of transgenic plants, particularly transgenic plants having enhanced disease resistance.

The present invention describes a novel pathogen-related nucleic acid and protein named PR1-C10. Nucleic acids involved in the disease resistance response in plants are needed for genetic manipulation of plants to exhibit specific phenotypic traits, particularly enhanced disease resistance. The present invention provides these and other advantages.

## SUMMARY OF THE INVENTION

Generally, it is the object of the present invention to provide nucleic acids and proteins relating to PR1-C10. It is an object of the present invention to provide transgenic plants comprising the nucleic acids of the present invention. It is another object of the present invention to provide methods for modulating, in a transgenic plant, the expression of the nucleic acids of the present invention.

Therefore, in one aspect, the present invention relates to an isolated nucleic acid comprising a member selected from the group consisting of (a) a polynucleotide encoding a polypeptide of the present invention; (b) a polynucleotide amplified from a *Zea mays* nucleic acid library using the primers of the present invention; (c) a polynucleotide comprising at least 20 contiguous bases of the polynucleotides of the present invention; (d) a polynucleotide encoding a maize PR1-C10 protein; (e) a polynucleotide having at least 85% sequence identity to the polynucleotides of the present invention; (f) a polynucleotide comprising at least 25 nucleotide in length which hybridizes under high stringency conditions to the polynucleotides of the present invention; (g) a polynucleotide of the present invention; and (h) a polynucleotide complementary to a polynucleotide of (a) through (f). The isolated nucleic acid can be DNA. The isolated nucleic acid can also be RNA.

In another aspect, the present invention relates to vectors comprising the polynucleotides of the present invention. Also the present invention relates to recombinant expression cassettes, comprising a nucleic acid of the present invention operably linked to a promoter.

In another aspect, the present invention is directed to a host cell into which has been introduced the recombinant expression cassette.

In yet another aspect, the present invention relates to a transgenic plant or plant cell comprising a recombinant expression cassette with a promoter operably linked to any  
5 of the isolated nucleic acids of the present invention. Preferred plants containing the recombinant expression cassette of the present invention include but are not limited to maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, and millet. The present invention also provides transgenic seed from the transgenic plant.

In another aspect, the present invention relates to an isolated protein selected from  
10 the group consisting of (a) a polypeptide comprising at least 25 contiguous amino acids of SEQ ID NO: 2; (b) a polypeptide which is a maize PR1-C10; (c) a polypeptide comprising at least 70% sequence identity to SEQ ID NO: 2; (d) a polypeptide encoded by a nucleic acid of the present invention; and (e) a polypeptide characterized by SEQ ID NO: 2.

15 In further aspect, the present invention relates to a method of modulating the level of protein in a plant by introducing into a plant cell a recombinant expression cassette comprising a polynucleotide of the present invention operably linked to a promoter; culturing the plant cell under plant growing conditions to produce a regenerated plant; and inducing expression of the polynucleotide for a time sufficient to modulate the  
20 protein of the present invention in the plant. Preferred plants of the present invention include but are not limited to maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, and millet. The level of protein in the plant can either be increased or decreased.

## 25 DETAILED DESCRIPTION OF THE INVENTION

### Overview

The present invention provides, among other things, compositions and methods for modulating (i.e., increasing or decreasing) the level of polynucleotides and



polypeptides of the present invention in plants. In particular, the polynucleotides and polypeptides of the present invention can be expressed temporally or spatially, e.g., at developmental stages, in tissues, and/or in quantities, which are uncharacteristic of non-recombinantly engineered plants. Thus, the present invention provides utility in such  
5 exemplary applications as engineering disease resistance in plants.

The present invention also provides isolated nucleic acid comprising polynucleotides of sufficient length and complementarity to a gene of the present invention to use as probes or amplification primers in the detection, quantitation, or isolation of gene transcripts. For example, isolated nucleic acids of the present invention  
10 can be used as probes in detecting deficiencies in the level of mRNA in screenings for desired transgenic plants, for detecting mutations in the gene (e.g., substitutions, deletions, or additions), for monitoring upregulation of expression or changes in enzyme activity in screening assays of compounds, for detection of any number of allelic variants (polymorphisms), orthologs, or paralogs of the gene, or for site directed mutagenesis in  
15 eukaryotic cells (see, e.g., U.S. patent No. 5,565,350). The isolated nucleic acids of the present invention can also be used for recombinant expression of their encoded polypeptides, or for use as immunogens in the preparation and/or screening of antibodies. The isolated nucleic acids of the present invention can also be employed for use in sense or antisense suppression of one or more genes of the present invention in a host cell,  
20 tissue, or plant. Attachment of chemical agents, which bind, intercalate, cleave and/or crosslink to the isolated nucleic acids of the present invention can also be used to modulate transcription or translation. The present invention also provides isolated proteins comprising a polypeptide of the present invention (e.g., preproenzyme, proenzyme, or enzymes).

25 The invention encompasses isolated or substantially purified nucleic acid or protein compositions. An "isolated" or "purified" nucleic acid molecule or protein, or biologically active portion thereof, is substantially or essentially free from components that normally accompany or interact with the nucleic acid molecule or protein as found in its naturally occurring environment. Thus, an isolated or purified nucleic acid molecule

or protein is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. Preferably, an "isolated" nucleic acid is free of sequences (preferably protein encoding sequences) that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, or 0.1 kb of nucleotide sequences that naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. A protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, 5%, or 1% (by dry weight) of contaminating protein. When the protein of the invention or biologically active portion thereof is recombinantly produced, preferably culture medium represents less than about 30%, 20%, 10%, 5%, or 1% (by dry weight) of chemical precursors or non-protein-of-interest chemicals.

The isolated nucleic acids and proteins of the present invention can be used over a broad range of plant types, particularly monocots such as the species of the family *Gramineae* including *Sorghum* (e.g. *S. bicolor*), *Triticum*, *Hordeum*, *Secale*, *Oryza*, and *Zea* (e.g. *Zea mays*). The isolated nucleic acid and proteins of the present invention can also be used in species from the genera: *Cucurbita*, *Rosa*, *Vitis*, *Juglans*, *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersicon*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Ciahorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Heterocallis*, *Nemesis*, *Pelargonium*, *Panieum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Pisum*, *Phaseolus*, *Lolium*, *Avena*, and *Allium*.

The present invention may be used for transformation of any plant species, including, but not limited to, monocots and dicots. Examples of plant species of interest include, but are not limited to, corn (*Zea mays*), *Brassica* sp. (e.g., *B. napus*, *B. rapa*, *B.*

*juncea*), particularly those *Brassica* species useful as sources of seed oil, alfalfa (*Medicago sativa*), rice (*Oryza sativa*), rye (*Secale cereale*), sorghum (*Sorghum bicolor*, *Sorghum vulgare*), millet (e.g., pearl millet (*Pennisetum glaucum*), proso millet (*Panicum miliaceum*), foxtail millet (*Setaria italica*), finger millet (*Eleusine coracana*)), sunflower (*Helianthus annuus*), safflower (*Carthamus tinctorius*), wheat (*Triticum aestivum*), soybean (*Glycine max*), tobacco (*Nicotiana tabacum*), potato (*Solanum tuberosum*), peanuts (*Arachis hypogaea*), cotton (*Gossypium barbadense*, *Gossypium hirsutum*), sweet potato (*Ipomoea batatas*), cassava (*Manihot esculenta*), coffee (*Coffea* spp.), coconut (*Cocos nucifera*), pineapple (*Ananas comosus*), citrus trees (*Citrus* spp.), cocoa (*Theobroma cacao*), tea (*Camellia sinensis*), banana (*Musa* spp.), avocado (*Persea americana*), fig (*Ficus casica*), guava (*Psidium guajava*), mango (*Mangifera indica*), olive (*Olea europaea*), papaya (*Carica papaya*), cashew (*Anacardium occidentale*), macadamia (*Macadamia integrifolia*), almond (*Prunus amygdalus*), sugar beets (*Beta vulgaris*), sugarcane (*Saccharum* spp.), oats, barley, vegetables, ornamentals, and conifers.

Vegetables include tomatoes (*Lycopersicon esculentum*), lettuce (e.g., *Lactuca sativa*), green beans (*Phaseolus vulgaris*), lima beans (*Phaseolus limensis*), peas (*Lathyrus* spp.), and members of the genus *Cucumis* such as cucumber (*C. sativus*), cantaloupe (*C. cantalupensis*), and musk melon (*C. melo*). Ornamentals include azalea (*Rhododendron* spp.), hydrangea (*Macrophylla hydrangea*), hibiscus (*Hibiscus rosasanensis*), roses (*Rosa* spp.), tulips (*Tulipa* spp.), daffodils (*Narcissus* spp.), petunias (*Petunia hybrida*), carnation (*Dianthus caryophyllus*), poinsettia (*Euphorbia pulcherrima*), and chrysanthemum.

Conifers that may be employed in practicing the present invention include, for example, pines such as loblolly pine (*Pinus taeda*), slash pine (*Pinus elliotii*), ponderosa pine (*Pinus ponderosa*), lodgepole pine (*Pinus contorta*), and Monterey pine (*Pinus radiata*); Douglas-fir (*Pseudotsuga menziesii*); Western hemlock (*Tsuga canadensis*); Sitka spruce (*Picea glauca*); redwood (*Sequoia sempervirens*); true firs such as silver fir (*Abies amabilis*) and balsam fir (*Abies balsamea*); and cedars such as Western red cedar (*Thuja plicata*) and Alaska yellow-cedar (*Chamaecyparis nootkatensis*). Preferably, plants of the present invention are crop plants (for example, corn, alfalfa, sunflower,

*Brassica*, soybean, cotton, safflower, peanut, sorghum, wheat, millet, tobacco, etc.), more preferably corn and soybean plants, yet more preferably corn plants.

Plants of particular interest include grain plants that provide seeds of interest, oil-seed plants, and leguminous plants. Seeds of interest include grain seeds, such as corn,  
5 wheat, barley, rice, sorghum, rye, etc. Oil-seed plants include cotton, soybean, safflower, sunflower, *Brassica*, maize, alfalfa, palm, coconut, etc. Leguminous plants include beans and peas. Beans include guar, locust bean, fenugreek, soybean, garden beans, cowpea, mungbean, lima bean, fava bean, lentils, chickpea, etc.

The invention is drawn to compositions and methods for inducing resistance in a  
10 plant to plant pests. Accordingly, the compositions and methods are also useful in protecting plants against fungal pathogens, viruses, nematodes, insects and the like.

By “disease resistance” is intended that the plants avoid the disease symptoms that are the outcome of plant-pathogen interactions. That is, pathogens are prevented from causing plant diseases and the associated disease symptoms, or alternatively, the  
15 disease symptoms caused by the pathogen is minimized or lessened.

By “antipathogenic compositions” is intended that the compositions of the invention have antipathogenic activity and thus are capable of suppressing, controlling, and/or killing the invading pathogenic organism. An antipathogenic composition of the invention will reduce the disease symptoms resulting from pathogen challenge by at least  
20 about 5% to about 50%, at least about 10% to about 60%, at least about 30% to about 70%, at least about 40% to about 80%, or at least about 50% to about 90% or greater. Hence, the methods of the invention can be utilized to protect plants from disease, particularly those diseases that are caused by plant pathogens.

Assays that measure antipathogenic activity are commonly known in the art, as  
25 are methods to quantitate disease resistance in plants following pathogen infection. See, for example, U.S. Patent No. 5,614,395, herein incorporated by reference. Such techniques include, measuring over time, the average lesion diameter, the pathogen biomass, and the overall percentage of decayed plant tissues. For example, a plant either expressing an antipathogenic polypeptide or having an antipathogenic composition

applied to its surface shows a decrease in tissue necrosis (*i.e.*, lesion diameter) or a decrease in plant death following pathogen challenge when compared to a control plant that was not exposed to the antipathogenic composition. Alternatively, antipathogenic activity can be measured by a decrease in pathogen biomass. For example, a plant  
5 expressing an antipathogenic polypeptide or exposed to an antipathogenic composition is challenged with a pathogen of interest. Over time, tissue samples from the pathogen-inoculated tissues are obtained and RNA is extracted. The percent of a specific pathogen RNA transcript relative to the level of a plant specific transcript allows the level of pathogen biomass to be determined. See, for example, Thomma *et al.* (1998) *Plant*  
10 *Biology* 95:15107-15111, herein incorporated by reference.

Furthermore, *in vitro* antipathogenic assays include, for example, the addition of varying concentrations of the antipathogenic composition to paper disks and placing the disks on agar containing a suspension of the pathogen of interest. Following incubation, clear inhibition zones develop around the discs that contain an effective concentration of  
15 the antipathogenic polypeptide (Liu *et al.* (1994) *Plant Biology* 91:1888-1892, herein incorporated by reference). Additionally, microspectrophotometrical analysis can be used to measure the *in vitro* antipathogenic properties of a composition (Hu *et al.* (1997) *Plant Mol. Biol.* 34:949-959 and Cammue *et al.* (1992) *J. Biol. Chem.* 267: 2228-2233, both of which are herein incorporated by reference).

20 Pathogens of the invention include, but are not limited to, viruses or viroids, bacteria, insects, nematodes, fungi, and the like. Viruses include any plant virus, for example, tobacco or cucumber mosaic virus, ringspot virus, necrosis virus, maize dwarf mosaic virus, etc. Specific fungal and viral pathogens for the major crops include:  
Soybeans: *Phytophthora megasperma* fsp. *glycinea*, *Macrophomina phaseolina*,  
25 *Rhizoctonia solani*, *Sclerotinia sclerotiorum*, *Fusarium oxysporum*, *Diaporthe phaseolorum* var. *sojae* (*Phomopsis sojae*), *Diaporthe phaseolorum* var. *caulivora*, *Sclerotium rolfsii*, *Cercospora kikuchii*, *Cercospora sojina*, *Peronospora manshurica*, *Colletotrichum dematium* (*Colletotrichum truncatum*), *Corynespora cassiicola*, *Septoria glycines*, *Phyllosticta sojicola*, *Alternaria alternata*, *Pseudomonas syringae* p.v. *glycinea*,

*Xanthomonas campestris* p.v. *phaseoli*, *Microsphaera diffusa*, *Fusarium semitectum*, *Phialophora gregata*, Soybean mosaic virus, *Glomerella glycines*, Tobacco Ring spot virus, Tobacco Streak virus, *Phakopsora pachyrhizi*, *Pythium aphanidermatum*, *Pythium ultimum*, *Pythium debaryanum*, Tomato spotted wilt virus, *Heterodera glycines* *Fusarium solani*; Canola: *Albugo candida*, *Alternaria brassicae*, *Leptosphaeria maculans*, *Rhizoctonia solani*, *Sclerotinia sclerotiorum*, *Mycosphaerella brassicicola*, *Pythium ultimum*, *Peronospora parasitica*, *Fusarium roseum*, *Alternaria alternata*; Alfalfa: *Clavibacter michiganese* subsp. *insidiosum*, *Pythium ultimum*, *Pythium irregulare*, *Pythium splendens*, *Pythium debaryanum*, *Pythium aphanidermatum*, *Phytophthora megasperma*, *Peronospora trifoliorum*, *Phoma medicaginis* var. *medicaginis*, *Cercospora medicaginis*, *Pseudopeziza medicaginis*, *Leptotrochila medicaginis*, *Fusarium*, *Xanthomonas campestris* p.v. *alfalfae*, *Aphanomyces euteiches*, *Stemphylium herbarum*, *Stemphylium alfalfae*; Wheat: *Pseudomonas syringae* p.v. *atrofaciens*, *Urocystis agropyri*, *Xanthomonas campestris* p.v. *translucens*, *Pseudomonas syringae* p.v. *syringae*, *Alternaria alternata*, *Cladosporium herbarum*, *Fusarium graminearum*, *Fusarium avenaceum*, *Fusarium culmorum*, *Ustilago tritici*, *Ascochyta tritici*, *Cephalosporium gramineum*, *Collotetrachium graminicola*, *Erysiphe graminis* f.sp. *tritici*, *Puccinia graminis* f.sp. *tritici*, *Puccinia recondita* f.sp. *tritici*, *Puccinia striiformis*, *Pyrenophora tritici-repentis*, *Septoria nodorum*, *Septoria tritici*, *Septoria avenae*, *Pseudocercospora herpotrichoides*, *Rhizoctonia solani*, *Rhizoctonia cerealis*, *Gaeumannomyces graminis* var. *tritici*, *Pythium aphanidermatum*, *Pythium arrhenomanes*, *Pythium ultimum*, *Bipolaris sorokiniana*, Barley Yellow Dwarf Virus, Brome Mosaic Virus, Soil Borne Wheat Mosaic Virus, Wheat Streak Mosaic Virus, Wheat Spindle Streak Virus, American Wheat Striate Virus, *Claviceps purpurea*, *Tilletia tritici*, *Tilletia laevis*, *Ustilago tritici*, *Tilletia indica*, *Rhizoctonia solani*, *Pythium arrhenomannes*, *Pythium graminicola*, *Pythium aphanidermatum*, High Plains Virus, European wheat striate virus; Sunflower: *Plasmophora halstedii*, *Sclerotinia sclerotiorum*, Aster Yellows, *Septoria helianthi*, *Phomopsis helianthi*, *Alternaria helianthi*, *Alternaria zinniae*, *Botrytis cinerea*, *Phoma macdonaldii*, *Macrophomina*

- phaseolina*, *Erysiphe cichoracearum*, *Rhizopus oryzae*, *Rhizopus arrhizus*, *Rhizopus stolonifer*, *Puccinia helianthi*, *Verticillium dahliae*, *Erwinia carotovorum* pv. *carotovora*, *Cephalosporium acremonium*, *Phytophthora cryptogea*, *Albugo tragopogonis*; Corn: *Fusarium moniliforme* var. *subglutinans*, *Erwinia stewartii*, *Fusarium moniliforme*,
- 5 *Gibberella zeae* (*Fusarium graminearum*), *Stenocarpella maydi* (*Diplodia maydis*), *Pythium irregulare*, *Pythium debaryanum*, *Pythium graminicola*, *Pythium splendens*, *Pythium ultimum*, *Pythium aphanidermatum*, *Aspergillus flavus*, *Bipolaris maydis* O, T (*Cochliobolus heterostrophus*), *Helminthosporium carbonum* I, II & III (*Cochliobolus carbonum*), *Exserohilum turcicum* I, II & III, *Helminthosporium pedicellatum*,
- 10 *Physoderma maydis*, *Phyllosticta maydis*, *Kabatiella maydis*, *Cercospora sorghi*, *Ustilago maydis*, *Puccinia sorghi*, *Puccinia polysora*, *Macrophomina phaseolina*, *Penicillium oxalicum*, *Nigrospora oryzae*, *Cladosporium herbarum*, *Curvularia lunata*, *Curvularia inaequalis*, *Curvularia pallescens*, *Clavibacter michiganense* subsp. *nebraskense*, *Trichoderma viride*, Maize Dwarf Mosaic Virus A & B, Wheat Streak
- 15 Mosaic Virus, Maize Chlorotic Dwarf Virus, *Claviceps sorghi*, *Pseudonomas avenae*, *Erwinia chrysanthemi* pv. *zea*, *Erwinia carotovora*, Corn stunt *spiroplasma*, *Diplodia macrospora*, *Sclerophthora macrospora*, *Peronosclerospora sorghi*, *Peronosclerospora philippinensis*, *Peronosclerospora maydis*, *Peronosclerospora sacchari*, *Sphacelotheca reiliana*, *Physopella zeae*, *Cephalosporium maydis*, *Cephalosporium acremonium*, Maize
- 20 Chlorotic Mottle Virus, High Plains Virus, Maize Mosaic Virus, Maize Rayado Fino Virus, Maize Streak Virus, Maize Stripe Virus, Maize Rough Dwarf Virus; Sorghum: *Exserohilum turcicum*, *Colletotrichum graminicola* (*Glomerella graminicola*), *Cercospora sorghi*, *Gloeocercospora sorghi*, *Ascochyta sorghina*, *Pseudomonas syringae* p.v. *syringae*, *Xanthomonas campestris* p.v. *holcicola*, *Pseudomonas andropogonis*,
- 25 *Puccinia purpurea*, *Macrophomina phaseolina*, *Perconia circinata*, *Fusarium moniliforme*, *Alternaria alternata*, *Bipolaris sorghicola*, *Helminthosporium sorghicola*, *Curvularia lunata*, *Phoma insidiosa*, *Pseudomonas avenae* (*Pseudomonas alboprecipitans*), *Ramulispora sorghi*, *Ramulispora sorghicola*, *Phyllachara sacchari*, *Sporisorium reilianum* (*Sphacelotheca reiliana*), *Sphacelotheca cruenta*, *Sporisorium*

*sorghi*, Sugarcane mosaic H, Maize Dwarf Mosaic Virus A & B, *Claviceps sorghi*,  
*Rhizoctonia solani*, *Acremonium strictum*, *Sclerophthora macrospora*,  
*Peronosclerospora sorghi*, *Peronosclerospora philippinensis*, *Sclerospora graminicola*,  
*Fusarium graminearum*, *Fusarium oxysporum*, *Pythium arrhenomanes*, *Pythium*  
5 *graminicola*, etc.

Nematodes include parasitic nematodes such as root-knot, cyst, and lesion  
nematodes, including *Heterodera* and *Globodera* spp; particularly *Globodera*  
*rostochiensis* and *globodera pailida* (potato cyst nematodes); *Heterodera glycines*  
(soybean cyst nematode); *Heterodera schachtii* (beet cyst nematode); and *Heterodera*  
10 *avenae* (cereal cyst nematode).

Insect pests include insects selected from the orders Coleoptera, Diptera,  
Hymenoptera, Lepidoptera, Mallophaga, Homoptera, Hemiptera, Orthoptera,  
Thysanoptera, Dermaptera, Isoptera, Anoplura, Siphonaptera, Trichoptera, etc.,  
particularly Coleoptera and Lepidoptera. Insect pests of the invention for the major crops  
15 include: Maize: *Ostrinia nubilalis*, European corn borer; *Agrotis ipsilon*, black cutworm;  
*Helicoverpa zea*, corn earworm; *Spodoptera frugiperda*, fall armyworm; *Diatraea*  
*grandiosella*, southwestern corn borer; *Elasmopalpus lignosellus*, lesser cornstalk borer;  
*Diatraea saccharalis*, sugarcane borer; *Diabrotica virgifera*, western corn rootworm;  
*Diabrotica longicornis barberi*, northern corn rootworm; *Diabrotica undecimpunctata*  
20 *howardi*, southern corn rootworm; *Melanotus* spp., wireworms; *Cyclocephala borealis*,  
northern masked chafer (white grub); *Cyclocephala immaculata*, southern masked chafer  
(white grub); *Popillia japonica*, Japanese beetle; *Chaetocnema pulicaria*, corn flea  
beetle; *Sphenophorus maidis*, maize billbug; *Rhopalosiphum maidis*, corn leaf aphid;  
*Anuraphis maidiradicis*, corn root aphid; *Blissus leucopterus leucopterus*, chinch bug;  
25 *Melanoplus femurrubrum*, redlegged grasshopper; *Melanoplus sanguinipes*, migratory  
grasshopper; *Hylemya platura*, seedcorn maggot; *Agromyza parvicornis*, corn blot  
leafminer; *Anaphothrips obscurus*, grass thrips; *Solenopsis milesta*, thief ant;  
*Tetranychus urticae*, twospotted spider mite; Sorghum: *Chilo partellus*, sorghum borer;  
*Spodoptera frugiperda*, fall armyworm; *Helicoverpa zea*, corn earworm; *Elasmopalpus*



*lignosellus*, lesser cornstalk borer; *Feltia subterranea*, granulate cutworm; *Phyllophaga crinita*, white grub; *Eleodes*, *Conoderus*, and *Aeolus spp.*, wireworms; *Oulema melanopus*, cereal leaf beetle; *Chaetocnema pulicaria*, corn flea beetle; *Sphenophorus maidis*, maize billbug; *Rhopalosiphum maidis*, corn leaf aphid; *Sipha flava*, yellow sugarcane aphid; *Blissus leucopterus leucopterus*, chinch bug; *Contarinia sorghicola*, sorghum midge; *Tetranychus cinnabarinus*, carmine spider mite; *Tetranychus urticae*, twospotted spider mite; Wheat: *Pseudaletia unipunctata*, army worm; *Spodoptera frugiperda*, fall armyworm; *Elasmopalpus lignosellus*, lesser cornstalk borer; *Agrotis orthogonia*, western cutworm; *Elasmopalpus lignosellus*, lesser cornstalk borer; *Oulema melanopus*, cereal leaf beetle; *Hypera punctata*, clover leaf weevil; *Diabrotica undecimpunctata howardi*, southern corn rootworm; Russian wheat aphid; *Schizaphis graminum*, greenbug; *Macrosiphum avenae*, English grain aphid; *Melanoplus femurrubrum*, redlegged grasshopper; *Melanoplus differentialis*, differential grasshopper; *Melanoplus sanguinipes*, migratory grasshopper; *Mayetiola destructor*, Hessian fly; *Sitodiplosis mosellana*, wheat midge; *Meromyza americana*, wheat stem maggot; *Hylemya coarctata*, wheat bulb fly; *Frankliniella fusca*, tobacco thrips; *Cephus cinctus*, wheat stem sawfly; *Aceria tulipae*, wheat curl mite; Sunflower: *Suleima helianthana*, sunflower bud moth; *Homoeosoma electellum*, sunflower moth; *zygogramma exclamationis*, sunflower beetle; *Bothyrus gibbosus*, carrot beetle; *Neolasioptera murtfeldtiana*, sunflower seed midge; Cotton: *Heliothis virescens*, cotton budworm; *Helicoverpa zea*, cotton bollworm; *Spodoptera exigua*, beet armyworm; *Pectinophora gossypiella*, pink bollworm; *Anthonomus grandis grandis*, boll weevil; *Aphis gossypii*, cotton aphid; *Pseudatomoscelis seriatus*, cotton fleahopper; *Trialeurodes abutilonea*, bandedwinged whitefly; *Lygus lineolaris*, tarnished plant bug; *Melanoplus femurrubrum*, redlegged grasshopper; *Melanoplus differentialis*, differential grasshopper; *Thrips tabaci*, onion thrips; *Frankliniella fusca*, tobacco thrips; *Tetranychus cinnabarinus*, carmine spider mite; *Tetranychus urticae*, twospotted spider mite; Rice: *Diatraea saccharalis*, sugarcane borer; *Spodoptera frugiperda*, fall armyworm; *Helicoverpa zea*, corn earworm; *Colaspis brunnea*, grape colaspis; *Lissorhoptrus oryzophilus*, rice water weevil;

*Sitophilus oryzae*, rice weevil; *Nephotettix nigropictus*, rice leafhopper; *Blissus leucopterus leucopterus*, chinch bug; *Acrosternum hilare*, green stink bug; Soybean: *Pseudoplusia includens*, soybean looper; *Anticarsia gemmatilis*, velvetbean caterpillar; *Plathypena scabra*, green cloverworm; *Ostrinia nubilalis*, European corn borer; *Agrotis* 5 *ipsilon*, black cutworm; *Spodoptera exigua*, beet armyworm; *Heliothis virescens*, cotton budworm; *Helicoverpa zea*, cotton bollworm; *Epilachna varivestis*, Mexican bean beetle; *Myzus persicae*, green peach aphid; *Empoasca fabae*, potato leafhopper; *Acrosternum hilare*, green stink bug; *Melanoplus femurrubrum*, redlegged grasshopper; *Melanoplus* 10 *differentialis*, differential grasshopper; *Hylemya platura*, seedcorn maggot; *Sericothrips variabilis*, soybean thrips; *Thrips tabaci*, onion thrips; *Tetranychus turkestani*, strawberry spider mite; *Tetranychus urticae*, twospotted spider mite; Barley: *Ostrinia nubilalis*, European corn borer; *Agrotis ipsilon*, black cutworm; *Schizaphis graminum*, greenbug; *Blissus leucopterus leucopterus*, chinch bug; *Acrosternum hilare*, green stink bug; *Euschistus servus*, brown stink bug; *Delia platura*, seedcorn maggot; *Mayetiola* 15 *destructor*, Hessian fly; *Petrobia latens*, brown wheat mite; Oil Seed Rape: *Brevicoryne brassicae*, cabbage aphid; *Phyllotreta cruciferae*, Flea beetle; *Mamestra configurata*, Bertha armyworm; *Plutella xylostella*, Diamond-back moth; *Delia* ssp., Root maggots.

Compositions of the invention include maize PR1-C10-like sequences that are involved in the pathogenic response of the plant. In particular, the present invention 20 provides for isolated nucleic acid molecules comprising nucleotide sequences encoding the amino acid sequences shown in SEQ ID NO:2, or the nucleotide sequences encoding the cDNA sequences of the plasmid deposited in a bacterial host as Patent Deposit No. PTA-1688. Further provided are polypeptides having an amino acid sequence encoded by a nucleic acid molecule described herein, for example those set forth in SEQ ID NO:1, 25 SEQ ID NO:3, or those deposited in a bacterial host as Patent Deposit No. PTA-1688, and fragments and variants thereof.

Plasmids containing the nucleotide sequences of the invention were deposited with the Patent Depository of the American Type Culture Collection (ATCC), Manassas, Virginia, on April 11, 2000 and assigned Patent Deposit No. PTA-1688. These deposits

will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. These deposits were made merely as a convenience for those of skill in the art and are not an admission that a deposit is required under 35 U.S.C. §112.

5

## Definitions

Units, prefixes, and symbols may be denoted in their SI accepted form. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation, amino acid sequences are written left to right in amino to carboxy orientation, respectively.

- 10 Numeric ranges are inclusive of the numbers defining the range and include each integer within the defined range. Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes. The terms defined below are
- 15 more fully defined by reference to the specification as a whole.

- By "amplified" is meant the construction of multiple copies of a nucleic acid sequence or multiple copies complementary to the nucleic acid sequence using at least one of the nucleic acid sequences as a template. Amplification systems include the polymerase chain reaction (PCR) system, ligase chain reaction (LCR) system, nucleic acid sequence based amplification (NASBA, Cangene, Mississauga, Ontario), Q-Beta Replicase systems, transcription-based amplification system (TAS), and strand displacement amplification (SDA). See, e.g., (Persing *et al.*) (1993) *Diagnostic Molecular Microbiology: Principles and Applications*, American Society for Microbiology, Washington, D.C. The product of amplification is termed an amplicon.
- 20

- 25 As used herein, "antisense orientation" includes reference to a duplex polynucleotide sequence, which is operably linked to a promoter in an orientation where the antisense strand is transcribed. The antisense strand is sufficiently complementary to an endogenous transcription product such that translation of the endogenous transcription product is often inhibited.

By “encoding” or “encoded”, with respect to a specified nucleic acid, is meant comprising the information for translation into the specified protein. A nucleic acid encoding a protein may comprise non-translated sequences (e.g., introns) within translated regions of the nucleic acid, or may lack such intervening non-translated sequences (e.g., as in cDNA). The information by which a protein is encoded is specified by the use of codons. Typically, the amino acid sequence is encoded by the nucleic acid using the “universal” genetic code. However, variants of the universal code, such as are present in some plant, animal, and fungal mitochondria, the bacterium *Mycoplasma capricolum*, or the ciliate *Macronucleus*, may be used when the nucleic acid is expressed therein.

When the nucleic acid is prepared or altered synthetically, advantage can be taken of known codon preferences of the intended host where the nucleic acid is to be expressed. For example, although nucleic acid sequences of the present invention may be expressed in both monocotyledonous and dicotyledonous plant species, sequences can be modified to account for the specific codon preferences and GC content preferences of monocotyledons or dicotyledons as these preferences have been shown to differ (Murray *et al.* (1989) *Nucl. Acids Res.* 17: 477-498). Thus, the maize preferred codon for a particular amino acid might be derived from known gene sequences from maize. Maize codon usage for 28 genes from maize plants is listed in Table 4 of Murray *et al.*, *supra*.

As used herein “full-length sequence” in reference to a specified polynucleotide or its encoded proteins means having the entire amino acid sequence of, a native (non-synthetic), endogenous, biologically active form of the specified protein. Methods to determine whether a sequence is full-length are well known in the art including such exemplary techniques as northern or western blots, primer extension, S1 protection, and ribonuclease protection. See, e.g., *Plant Molecular Biology: A Laboratory Manual*, (1997) Clark, Springer-Verlag, Berlin. Comparison to known full-length homologous (orthologous and/or paralogous) sequences can also be used to identify full-length sequences of the present invention. Additionally, consensus sequences typically present at the 5’ and 3’ untranslated regions of mRNA aid in the identification of a

polynucleotide as full-length. For example, the consensus sequence, ANNNNAUGG, where the underlined codon represents the N-terminal methionine, aids in determining whether the polynucleotide has a complete 5' end. Consensus sequences at the 3' end, such as polyadenylation sequences, aid in determining whether the polynucleotide has a complete 3' end.

As used herein, "heterologous" in reference to a nucleic acid is a nucleic acid that originates from a foreign species, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. For example, a promoter operably linked to a heterologous structural gene is from a species different from that from which the structural gene was derived, or, if from the same species, one or both are substantially modified from their original form. A heterologous protein may originate from a foreign species, or, if from the same species, is substantially modified from its original form by deliberate human intervention.

By "host cell" is meant a cell, which contains a vector and supports the replication and/or expression of the vector. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, insect, amphibian, or mammalian cells. Preferably, host cells are monocotyledonous or dicotyledonous plant cells. A particularly preferred monocotyledonous host cell is a maize host cell.

The term "introduced" in the context of inserting a nucleic acid into a cell, means "transfection" or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid into a eukaryotic or prokaryotic cell where the nucleic acid may be incorporated into the genome of the cell (e.g., chromosome, plasmid, plastid or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (e.g., transfected mRNA).

The terms "isolated" refers to material, such as a nucleic acid or a protein, which is: (1) substantially or essentially free from components that normally accompany or interact with it as found in its naturally occurring environment. The isolated material optionally comprises material not found with the material in its natural environment; or (2) if the material is in its natural environment, the material has been synthetically (non-

naturally) altered by deliberate human intervention to a composition and/or placed at a location in the cell (e.g., genome or subcellular organelle) not native to a material found in that environment. The alteration to yield the synthetic material can be performed on the material within or removed from its natural state. For example, a naturally occurring

5 nucleic acid becomes an isolated nucleic acid if it is altered, or if it is transcribed from DNA which has been altered, by means of human intervention performed within the cell from which it originates. See, e.g., Compounds and Methods for Site Directed Mutagenesis in Eukaryotic Cells, Kmiec, U.S. patent No. 5,565,350; *In Vivo* Homologous Sequence Targeting in Eukaryotic Cells, Zarling *et al.*, PCT/US93/03868. Likewise, a

10 naturally occurring nucleic acid (e.g., a promoter) becomes isolated if it is introduced by non-naturally occurring means to a locus of the genome not native to that nucleic acid. Nucleic acids, which are “isolated”, as defined herein, are also referred to as “heterologous” nucleic acids.

As used herein, “nucleic acid” includes reference to a deoxyribonucleotide or

15 ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues having the essential nature of natural nucleotides in that they hybridize to single-stranded nucleic acids in a manner similar to naturally occurring nucleotides (e.g., peptide nucleic acids).

By “nucleic acid library” is meant a collection of isolated DNA or RNA

20 molecules, which comprise and substantially represent the entire transcribed fraction of a genome of a specified organism. Construction of exemplary nucleic acid libraries, such as genomic and cDNA libraries, is taught in standard molecular biology references such as Berger and Kimmel (1989) *Guide to Molecular Cloning Techniques, Methods in Enzymology*, Vol. 152, Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.*,

25 *Molecular Cloning – A Laboratory Manual*, 2<sup>nd</sup> ed., Vol. 1-3; and Ausubel *et al.* (1994) *Current Protocols in Molecular Biology*, Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc.

As used herein “operably linked” includes reference to a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates and

mediates transcription of the DNA sequence corresponding to the second sequence. Generally, operably linked means that the nucleic acid sequences being linked are contiguous and, where necessary to join two protein coding regions, contiguous and in the same reading frame.

5 As used herein, the term “plant” includes reference to whole plants, plant organs (e.g., leaves, stems, roots, etc.), seeds and plant cells and progeny of same. Plant cell, as used herein includes, without limitation, seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores. The class of plants, which can be used in the methods of the  
10 invention, is generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants. Preferred plants include, but are not limited to maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, and millet. A particularly preferred plant is maize (*Zea mays*).

15 As used herein, “polynucleotide” includes reference to a deoxyribopolynucleotide, ribopolynucleotide, or analogs thereof that have the essential nature of a natural ribonucleotide in that they hybridize, under stringent hybridization conditions, to substantially the same nucleotide sequence as naturally occurring nucleotides and/or allow translation into the same amino acid(s) as the naturally  
20 occurring nucleotide(s). A polynucleotide can be full-length or a subsequence of a native or heterologous structural or regulatory gene. Unless otherwise indicated, the term includes reference to the specified sequence as well as the complementary sequence thereof. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are “polynucleotides” as that term is intended herein. Moreover, DNAs or RNAs  
25 comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modification have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide as it is employed herein embraces such chemically, enzymatically or metabolically modified

forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including among other things, simple and complex cells.

The terms “polypeptide”, “peptide” and “protein” are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. The essential nature of such analogues of naturally occurring amino acids is that, when incorporated into a protein, that protein is specifically reactive to antibodies elicited to the same protein but consisting entirely of naturally occurring amino acids.

The terms “polypeptide”, “peptide”, and “protein” are also inclusive of modifications including, but not limited to, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation. It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of post-translation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Further, this invention contemplates the use of both the methionine containing and the methionine-less amino terminal variants of the protein of the invention.

As used herein “promoter” includes reference to a region of DNA upstream from the start of transcription and involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. A “plant promoter” is a promoter capable of initiating transcription in plant cells whether or not its origin is a plant cell. Exemplary plant promoters include, but are not limited to, those that are obtained from plants, plant viruses, and bacteria which comprise genes expressed in plant cells such as *Agrobacterium* or *Rhizobium*. Examples of promoters under developmental control include promoters that preferentially initiate transcription in certain tissues, such as



leaves, roots, or seeds. Such promoters are referred to as “tissue preferred”. A “cell type” preferred promoter primarily drives expression in certain cell types in one of more organs, for example, vascular cells in roots and leaves. An “inducible” or “repressible” promoter is a promoter, which is under environmental control. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions or the presence of light. Tissue preferred, cell type preferred and inducible promoters constitute the class of “non-constitutive” promoters. A “constitutive” promoter is a promoter, which is active under most environmental conditions.

As used herein “recombinant” includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under-expressed or not expressed at all as a result of deliberate human intervention. The term “recombinant” as used herein does not encompass the alteration of the cell or vector by naturally occurring events (e.g., spontaneous mutation, natural transformation/transduction/transposition) such as those occurring without deliberate human intervention.

As used herein, a “recombinant expression cassette” is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements, which permit transcription of a particular nucleic acid in a host cell. The recombinant expression cassette can be incorporated into a plasmid, chromosome, mitochondrial DNA, plastid DNA, virus, or nucleic acid fragment. Typically, the recombinant expression cassette portion of an expression vector includes, among other sequences, a nucleic acid to be transcribed, and a promoter.

The term “residue” or “amino acid residue” or “amino acid” are used interchangeably herein to refer to an amino acid that is incorporated into a protein, polypeptide, or peptide (collectively “protein”). The amino acid may be a naturally occurring amino acid and, unless otherwise limited, may encompass non-natural analogs

of natural amino acids that can function in a similar manner as naturally occurring amino acids.

5 The term “selectively hybridizes” includes a reference to hybridization, under stringent hybridization conditions, of a nucleic acid sequence to a specified nucleic acid target sequence to a detectably greater degree (e.g., at least 2-fold over background) than its hybridization to non-target nucleic acid sequences and to the substantial exclusion of non-target nucleic acids.

10 The terms “stringent conditions” or “stringent hybridization conditions” include reference to conditions under which a probe will hybridize to its target sequence, to a detectably greater degree than other sequences (e.g., at least 2-fold over background). Stringent conditions are sequence-dependent and will be different in different circumstances. By controlling the stringency of the hybridization and/or washing conditions, target sequences can be identified which are 100% complementary to the probe (homologous probing). Alternatively, stringency conditions can be adjusted to  
15 allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous probing). Generally, a probe is less than about 1000 nucleotides in length, optionally less than 500 nucleotides in length.

Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other  
20 salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS  
25 (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC

at 60 to 65°C. Optionally, wash buffers may comprise about 0.1% to about 1% SDS. Duration of hybridization is generally less than about 24 hours, usually about 4 to about 12.

Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the  $T_m$  can be approximated from the equation of Meinkoth and Wahl (1984) *Anal. Biochem.*, 138:267-284:  $T_m = 81.5^\circ\text{C} + 16.6 (\log M) + 0.41 (\%CG) - 0.61 (\% \text{ form}) - 500/L$ ; where M is the molarity of monovalent cations, %CG is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe.  $T_m$  is reduced by about 1°C for each 1% of mismatching; thus,  $T_m$ , hybridization and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with  $\geq 90\%$  identity are sought, the  $T_m$  can be decreased 10°C. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point ( $T_m$ ) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4° lower than the thermal melting point ( $T_m$ ); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the thermal melting point ( $T_m$ ); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20°C lower than the thermal melting point ( $T_m$ ). Using the equation, hybridization and wash compositions, and desired  $T_m$ , those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a  $T_m$  of less than 45°C (aqueous solution) or 32°C (formamide solution) it is preferred to increase the SSC concentration so that a higher temperature can be used. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes*, Part I,

Chapter 2 “Overview of principles of hybridization and the strategy of nucleic acid probe assays”, Elsevier, New York; and Ausubel, *et al.* (1995) *Current Protocols in Molecular Biology*, Chapter 2, Greene Publishing and Wiley-Interscience, New York.

As used herein, “transgenic plant” includes reference to a plant, which comprises within its genome a heterologous polynucleotide. Generally, the heterologous polynucleotide is stably integrated within the genome such that the polynucleotide is passed on to successive generations. The heterologous polynucleotide may be integrated into the genome alone or as part of a recombinant expression cassette. “Transgenic” is used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses or asexual propagation from the initial transgenic. The term “transgenic” as used herein does not encompass the alteration of the genome (chromosomal or extra-chromosomal) by conventional plant breeding methods or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation.

As used herein, “vector” includes reference to a nucleic acid used in transfection of a host cell and into which can be inserted a polynucleotide. Vectors are often replicons. Expression vectors permit transcription of a nucleic acid inserted therein.

The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) “reference sequence”, (b) “comparison window”, (c) “sequence identity”, (d) “percentage of sequence identity”, and (e) “substantial identity”.

(a) As used herein, “reference sequence” is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence.

(b) As used herein, “comparison window” makes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence in

the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well known in the art. Thus, the determination of percent sequence identity between any two sequences can be accomplished using a mathematical algorithm. Non-limiting examples of such mathematical algorithms are the algorithm of Myers and Miller (1988) *CABIOS* 4:11-17; the local homology algorithm of Smith *et al.* (1981) *Adv. Appl. Math.* 2:482; the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443-453; the search-for-similarity-method of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci.* 85:2444-2448; the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877.

Computer implementations of these mathematical algorithms can be utilized for comparison of sequences to determine sequence identity. Such implementations include, but are not limited to: CLUSTAL in the PC/Gene program (available from Intelligenetics, Mountain View, California); the ALIGN program (Version 2.0) and GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Version 8 (available from Genetics Computer Group (GCG), 575 Science Drive, Madison, Wisconsin, USA). Alignments using these programs can be performed using the default parameters. The CLUSTAL program is well described by Higgins *et al.* (1988) *Gene* 73:237-244 (1988); Higgins *et al.* (1989) *CABIOS* 5:151-153; Corpet *et al.* (1988) *Nucleic Acids Res.* 16:10881-90; Huang *et al.* (1992) *CABIOS* 8:155-65; and Pearson *et al.* (1994) *Meth. Mol. Biol.* 24:307-331. The ALIGN program is based on the algorithm of Myers and Miller (1988) *supra*. A PAM120 weight residue table, a gap

length penalty of 12, and a gap penalty of 4 can be used with the ALIGN program when comparing amino acid sequences. The BLAST programs of Altschul *et al* (1990) *J. Mol. Biol.* 215:403 are based on the algorithm of Karlin and Altschul (1990) *supra*. BLAST nucleotide searches can be performed with the BLASTN program, score = 100,

5 wordlength = 12, to obtain nucleotide sequences homologous to a nucleotide sequence encoding a protein of the invention. BLAST protein searches can be performed with the BLASTX program, score = 50, wordlength = 3, to obtain amino acid sequences homologous to a protein or polypeptide of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST (in BLAST 2.0) can be utilized as described in  
10 Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389. Alternatively, PSI-BLAST (in BLAST 2.0) can be used to perform an iterated search that detects distant relationships between molecules. See Altschul *et al.* (1997) *supra*. When utilizing BLAST, Gapped BLAST, PSI-BLAST, the default parameters of the respective programs (e.g., BLASTN for nucleotide sequences, BLASTX for proteins) can be used. See

15 <http://www.ncbi.nlm.nih.gov>. Alignment may also be performed manually by inspection.

Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using GAP version 10 using the following parameters: % identity using GAP Weight of 50 and Length Weight of 3; % similarity using Gap Weight of 12 and Length Weight of 4, or any equivalent program. By "equivalent program" is  
20 intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by the preferred program.

GAP uses the algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48: 443-  
25 453, to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. GAP considers all possible alignments and gap positions and creates the alignment with the largest number of matched bases and the fewest gaps. It allows for the provision of a gap creation penalty and a gap extension penalty in units of matched bases. GAP must make a profit of gap creation penalty

number of matches for each gap it inserts. If a gap extension penalty greater than zero is chosen, GAP must, in addition, make a profit for each gap inserted of the length of the gap times the gap extension penalty. Default gap creation penalty values and gap extension penalty values in Version 10 of the Wisconsin Genetics Software Package for protein sequences are 8 and 2, respectively. For nucleotide sequences the default gap creation penalty is 50 while the default gap extension penalty is 3. The gap creation and gap extension penalties can be expressed as an integer selected from the group of integers consisting of from 0 to 200. Thus, for example, the gap creation and gap extension penalties can be 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65 or greater.

GAP presents one member of the family of best alignments. There may be many members of this family, but no other member has a better quality. GAP displays four figures of merit for alignments: Quality, Ratio, Identity, and Similarity. The Quality is the metric maximized in order to align the sequences. Ratio is the quality divided by the number of bases in the shorter segment. Percent Identity is the percent of the symbols that actually match. Percent Similarity is the percent of the symbols that are similar. Symbols that are across from gaps are ignored. A similarity is scored when the scoring matrix value for a pair of symbols is greater than or equal to 0.50, the similarity threshold. The scoring matrix used in Version 10 of the Wisconsin Genetics Software Package is BLOSUM62 (*see* Henikoff and Henikoff, *Proc Natl Acad Sci USA* 89:10915).

Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using the BLAST 2.0 suite of programs using default parameters. Altschul *et al.*, (1997) *Nucleic Acids Res.* 25:3389-3402 or GAP version 10 of Wisconsin Genetic Software Package using default parameters. Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the

neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for  
5 nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always <0) . For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or  
10 below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid  
15 sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89: 10915).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.,* Karlin &  
20 Altschul, (1993) *Proc. Nat'l. Acad. Sci. USA* 90:5873-5877). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability that a match between two nucleotide or two amino acid sequences would occur by chance.

BLAST searches assume that proteins can be modeled as random sequences.  
25 However, many real proteins comprise regions of nonrandom sequences, which may be homopolymeric tracts, short-period repeats, or regions enriched in one or more amino acids. Such low-complexity regions may be aligned between unrelated proteins even though other regions of the protein are entirely dissimilar. A number of low-complexity filter programs can be employed to reduce such low-complexity alignments. For



example, the SEG (Wooten and Federhen (1993) *Comput. Chem.*, 17:149-163) and XNU (Claverie and States (1993) *Comput. Chem.*, 17:191-201) low-complexity filters can be employed alone or in combination.

(c) As used herein, “sequence identity” or “identity” in the context of two nucleic acid or polypeptide sequences makes reference to the residues in the two sequences, which are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g. charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences, which differ by such conservative substitutions, are said to have “sequence similarity” or “similarity”. Means for making this adjustment are well known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., according to the algorithm of Meyers and Miller, (1998) *Computer Applic. Biol. Sci.*, 4: 11-17 e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

(d) As used herein, “percentage of sequence identity” means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of

matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

(e)(i) The term “substantial identity” of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 65%, 70% sequence identity to a reference sequence, preferably at least 80%, more preferably at least 90%, 91%, 92%, 93%, 94%, yet more preferably 95%, 96%, 97%, 98%, or 99% sequence identity compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill in the art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning, and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 60%, more preferably at least 70%, 80%, 90%, and most preferably at least 95%.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. However, stringent conditions encompass temperatures in the range of about 1°C to about 20°C lower than the  $T_m$ , depending upon the desired degree of stringency as otherwise qualified herein. Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides they encode are substantially identical. This may occur, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. One indication that two nucleic acid sequences are substantially identical is when the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

(e)(ii) The term “substantial identity” in the context of a peptide indicates that a peptide comprises a sequence with at least 70% sequence identity to a reference sequence, preferably 80%, more preferably 85%, most preferably at least 90% or 95%

sequence identity to the reference sequence over a specified comparison window.

Preferably, optimal alignment is conducted using the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443-453. An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution. Peptides that are “substantially similar” share sequences as noted above except that residue positions that are not identical may differ by conservative amino acid changes.

#### Nucleic Acids

The present invention provides, among other things, isolated nucleic acids of RNA, DNA, and analogs and/or chimeras thereof, comprising a polynucleotide of the present invention.

A polynucleotide of the present invention is inclusive of:

(a) a polynucleotide encoding a polypeptide of SEQ ID NO:2, including the exemplary polynucleotides of SEQ ID NO:1 and SEQ ID NO:3;

(b) a polynucleotide which is the product of amplification from a *Zea mays* nucleic acid library using primer pairs which selectively hybridize under stringent conditions to loci within a polynucleotide selected from the group consisting of SEQ ID NO:1, wherein the polynucleotide has substantial sequence identity to a polynucleotide selected from the group consisting of SEQ ID NO:1;

(c) a polynucleotide which selectively hybridizes to a polynucleotide of (a) or (b);

(d) a polynucleotide having a specified sequence identity with the polynucleotide of SEQ ID NO:1;

(e) a polynucleotide comprising at least 20 contiguous bases of SEQ ID NO:1;

and

(e) complementary sequences of polynucleotides of (a), (b), (c), or (d).

*A. Polynucleotides Encoding a Polypeptide of the Present Invention*

The present invention provides isolated nucleic acids comprising a polynucleotide of the present invention, wherein the polynucleotide encodes a polypeptide of the present invention, or conservatively modified or polymorphic variants thereof. Accordingly, the present invention includes polynucleotides of SEQ ID NO:1, nucleotides 63-674 of SEQ ID NO:1 (SEQ ID NO:3), and silent variations of polynucleotides encoding a polypeptide of SEQ ID NO:2. Additionally, the present invention further provides isolated nucleic acids comprising polynucleotides encoding one or more allelic (polymorphic) variants of polypeptides/polynucleotides. Polymorphic variants are frequently used to follow segregation of chromosomal regions in, for example, marker assisted selection methods for crop improvement.

Thus, the genes and nucleotide sequences of the invention include both the naturally occurring sequences as well as mutant forms. Likewise, the proteins of the invention encompass both naturally occurring proteins as well as variations and modified forms thereof. Such variants will continue to possess the desired pathogenesis related activity. Obviously, the mutations that will be made in the DNA encoding the variant must not place the sequence out of reading frame and preferably will not create complementary regions that could produce secondary mRNA structure. See, EP Patent Application Publication No. 75,444.

The deletions, insertions, and substitutions of the polypeptide sequences encompassed herein are not expected to produce radical changes in the characteristics of the protein. However, when it is difficult to predict the exact effect of the substitution, deletion, or insertion in advance of doing so, one skilled in the art will appreciate that the effect will be evaluated by routine screening assays. That is, the activity can be evaluated by Northern blots. See, for example, Gordon-Weeks, *et al.* (1997) *Physiol. Mol. Plant Pathol.* 50:263-273, herein incorporated by reference.

Variant nucleotide sequences and proteins also encompass sequences and proteins derived from a mutagenic and recombinogenic procedure such as DNA shuffling. With such a procedure, one or more different PR1-C10 coding sequences can be manipulated

to create a new PR1-C10 possessing the desired properties. In this manner, libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides comprising sequence regions that have substantial sequence identity and can be homologously recombined *in vitro* or *in vivo*. For example, using this approach,

5 sequence motifs encoding a domain of interest may be shuffled between the PR1-C10 gene of the invention and other known PR1 genes to obtain a new gene coding for a protein with an improved property of interest, such as an increased  $K_m$  in the case of an enzyme. Strategies for such DNA shuffling are known in the art. See, for example, Stemmer (1994) *Proc. Natl. Acad. Sci. USA* 91:10747-10751; Stemmer (1994) *Nature*

10 370:389-391; Cramer *et al.* (1997) *Nature Biotech.* 15:436-438; Moore *et al.* (1997) *J. Mol. Biol.* 272:336-347; Zhang *et al.* (1997) *Proc. Natl. Acad. Sci. USA* 94:4504-4509; Cramer *et al.* (1998) *Nature* 391:288-291; and U.S. Patent Nos. 5,605,793 and 5,837,458.

#### 15 B. Polynucleotides Amplified from a *Zea mays* Nucleic Acid Library

The present invention provides an isolated nucleic acid comprising a polynucleotide of the present invention, wherein the polynucleotides are amplified from a *Zea mays* nucleic acid library. *Zea mays* lines B73, PHRE1, A632, BMS-P2#10, W23, and Mo17 are known and publicly available. Other publicly known and available maize

20 lines can be obtained from the Maize Genetics Cooperation (Urbana, IL). The nucleic acid library may be a cDNA library, a genomic library, or a library generally constructed from nuclear transcripts at any stage of intron processing. cDNA libraries can be normalized to increase the representation of relatively rare cDNAs. In optional embodiments, the cDNA library is constructed using a full-length cDNA synthesis

25 method. Examples of such methods include Oligo-Capping (Maruyama, K. and Sugano, S. *Gene* 138: 171-174, 1994), Biotinylated CAP Trapper (Carninci, P., Kvan, C., *et al.* *Genomics* 37: 327-336, 1996), and CAP Retention Procedure (Edery, E., Chu, L.L., *et al.* *Molecular and Cellular Biology* 15: 3363-3371, 1995). cDNA synthesis is often catalyzed at 50-55°C to prevent formation of RNA secondary structure. Examples of

reverse transcriptases that are relatively stable at these temperatures are SuperScript II Reverse Transcriptase (Life Technologies, Inc.) AMV Reverse Transcriptase (Boehringer Mannheim) and RetroAmp Reverse Transcriptase (Epicentre). Rapidly growing tissues, or rapidly dividing cells are preferably used as mRNA sources.

5           The present invention also provides subsequences of the polynucleotides of the present invention. A variety of subsequences can be obtained using primers which selectively hybridize under stringent conditions to at least two sites within a polynucleotide of the present invention, or to two sites within the nucleic acid which flank and comprise a polynucleotide of the present invention, or to a site within a  
10 polynucleotide of the present invention and a site within the nucleic acid which comprises it. Primers are chosen to selectively hybridize, under stringent hybridization conditions, to a polynucleotide of the present invention. Generally, the primers are complementary to a subsequence of the target nucleic acid, which they amplify. As those skilled in the art will appreciate, the sites to which the primer pairs will selectively  
15 hybridize are chosen such that a single contiguous nucleic acid can be formed under the desired amplification conditions. In optional embodiments, the primer will be constructed so that they selectively hybridize under stringent conditions to a sequence (or its complement) within the target nucleic acid which comprises the codon encoding the carboxy or amino terminal amino acid residue (i.e., the 3' terminal coding region and 5'  
20 terminal coding region, respectively) of the polynucleotides of the present invention. Optionally within these embodiments, the primers will be constructed to selectively hybridize entirely within the coding region of the target polynucleotide of the present invention such that the product of amplification of a cDNA target will consist of the coding region of that cDNA. The primer length in nucleotides is selected from the group  
25 of integers consisting of from at least 15 to 50. Thus, the primers can be at least 15, 18, 20, 25, 30, 40, or 50 nucleotides in length. Those of skill will recognize that a lengthened primer sequence can be employed to increase specificity of binding (i.e., annealing) to a target sequence. A non-annealing sequence at the 5' end of a primer (a "tail") can be added, for example, to introduce a cloning site at the terminal ends of the amplicon.

The amplification products can be translated using expression systems well known to those of skill in the art and as discussed, *infra*. The resulting translation products can be confirmed as polypeptides of the present invention by, for example, assaying for the appropriate catalytic activity (e.g., specific activity and/or substrate specificity), or verifying the presence of one or more linear epitopes, which are specific to a polypeptide of the present invention. Methods for protein synthesis from PCR derived templates are known in the art and available commercially. See, e.g., Amersham Life Sciences, Inc, Catalog '97, p. 354.

Methods for obtaining 5' and/or 3' ends of a vector insert are well known in the art. See, e.g., RACE (Rapid Amplification of Complementary Ends) as described in Frohman, (1990) in *PCR Protocols: A Guide to Methods and Applications*, M. A. Innis, D. H. Gelfand, J. J. Sninsky, T. J. White, Academic Press, Inc., San Diego, pp. 28-38); see also U.S. Pat. No. 5,470,722, and Ausubel, *et al.* (1995) *Current Protocols in Molecular Biology*, Unit 15.6, Greene Publishing and Wiley-Interscience, New York; and Frohman and Martin (1989) *Techniques* 1:165.

#### *C. Polynucleotides Which Selectively Hybridize to a Polynucleotide of (A) or (B)*

The present invention provides isolated nucleic acids comprising polynucleotides of the present invention, wherein the polynucleotides selectively hybridize, under selective hybridization conditions, to a polynucleotide of section (A) or (B) as discussed above. Thus, the polynucleotides of this embodiment can be used for isolating, detecting, and/or quantifying nucleic acids comprising the polynucleotides of (A) or (B). For example, polynucleotides of the present invention can be used to identify, isolate, or amplify partial or full-length clones in a deposited library. In some embodiments, the polynucleotides are genomic or cDNA sequences isolated or otherwise complementary to a cDNA from a dicot or monocot nucleic acid library. Exemplary species of monocots and dicots include, but are not limited to: corn, canola, soybean, cotton, wheat, sorghum, sunflower, oats, sugar cane, millet, barley, and rice. Optionally, the cDNA library comprises at least 80% full-length sequences, preferably at least 85% or 90% full-length

sequences, and more preferably at least 95% full-length sequences. The cDNA libraries can be normalized to increase the representation of rare sequences. Low stringency hybridization conditions are typically, but not exclusively, employed with sequences having a reduced sequence identity relative to complementary sequences. Moderate and high stringency conditions can optionally be employed for sequences of greater identity. Low stringency conditions allow selective hybridization of sequences having about 70% sequence identity and can be employed to identify orthologous or paralogous sequences. By "orthologs" is intended genes derived from a common ancestral gene and which are found in different species as a result of speciation. Genes found in different species are considered orthologs when their nucleotide sequences and/or their encoded protein sequences share substantial identity as defined elsewhere herein. Functions of orthologs are often highly conserved among species.

In this manner, methods such as PCR, hybridization, and the like can be used to identify such sequences based on their sequence homology to the sequences set forth herein. Sequences isolated based on their sequence identity to the entire PR1-C10 sequences set forth herein or to fragments thereof are encompassed by the present invention. Such sequences include sequences that are orthologs of the disclosed sequences. Thus, isolated sequences that encode for a PRI-C10 protein having biological activity and which hybridize under stringent conditions to the sequences disclosed herein, or to fragments thereof, are encompassed by the present invention.

In a PCR approach, oligonucleotide primers can be designed for use in PCR reactions to amplify corresponding DNA sequences from cDNA or genomic DNA extracted from any plant of interest, particularly maize. Methods for designing PCR primers and PCR cloning are generally known in the art and are disclosed in Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York). See also Innis *et al.*, eds. (1990) *PCR Protocols: A Guide to Methods and Applications* (Academic Press, New York); Innis and Gelfand, eds. (1995) *PCR Strategies* (Academic Press, New York); and Innis and Gelfand, eds. (1999) *PCR Methods Manual* (Academic Press, New York). Known



methods of PCR include, but are not limited to, methods using paired primers, nested primers, single specific primers, degenerate primers, gene-specific primers, vector-specific primers, partially-mismatched primers, and the like.

In hybridization techniques, all or part of a known nucleotide sequence is used as a probe that selectively hybridizes to other corresponding nucleotide sequences present in a population of cloned genomic DNA fragments or cDNA fragments (i.e., genomic or cDNA libraries) from a chosen organism. The hybridization probes may be genomic DNA fragments, cDNA fragments, RNA fragments, or other oligonucleotides, and may be labeled with a detectable group such as <sup>32</sup>P, or any other detectable marker. Thus, for example, probes for hybridization can be made by labeling synthetic oligonucleotides based on the PR1-C10 sequences of the invention. Methods for preparation of probes for hybridization and for construction of cDNA and genomic libraries are generally known in the art and are disclosed in Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York).

For example, the entire PR1-C10 sequence disclosed herein, or one or more portions thereof, may be used as a probe capable of specifically hybridizing to corresponding PR1-C10 sequences and messenger RNAs. To achieve specific hybridization under a variety of conditions, such probes include sequences that are unique among PR1-C10 sequences and are preferably at least about 10 nucleotides in length, and most preferably at least about 20 nucleotides in length. Such probes may be used to amplify corresponding PR1-C10 sequences from a chosen plant by PCR. This technique may be used to isolate additional coding sequences from a desired plant or as a diagnostic assay to determine the presence of coding sequences in a plant. Hybridization techniques include hybridization screening of plated DNA libraries (either plaques or colonies; see, for example, Sambrook *et al.* (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York).

*D. Polynucleotides Having a Specific Sequence Identity with the Polynucleotides of (A), (B) or (C)*

The present invention provides isolated nucleic acids comprising polynucleotides of the present invention, wherein the polynucleotides have a specified identity at the  
5 nucleotide level to a polynucleotide as disclosed above in sections (A), (B), or (C). The percentage of identity to a reference sequence is at least 60% and, rounded upwards to the nearest integer, can be expressed as an integer selected from the group of integers consisting of from 60 to 99. Thus, for example, the percentage of identity to a reference  
10 sequence can be at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater.

*E. Polynucleotides Complementary to the Polynucleotides of (A)-(D).*

The present invention provides isolated nucleic acids comprising polynucleotides complementary to the polynucleotides of sections (A)-(D), above. As those of skill in the  
15 art will recognize, complementary sequences base-pair throughout the entirety of their length with the polynucleotides of sections (A)-(D) (i.e., have 100% sequence identity over their entire length). Complementary bases associate through hydrogen bonding in double stranded nucleic acids. For example, the following base pairs are complementary: guanine and cytosine; adenine and thymine; and adenine and uracil.  
20

*F. Polynucleotides Which are Subsequences of the Polynucleotides of The Present Invention*

The present invention provides isolated nucleic acids comprising polynucleotides which comprise at least 20 contiguous bases from the polynucleotide characterized in  
25 SEQ ID NO:1. The length of the polynucleotide is given as an integer selected from the group consisting of from at least 15 to the length of the nucleic acid sequence from which the polynucleotide is a subsequence of. Thus, for example, polynucleotides of the present invention are inclusive of polynucleotides comprising at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45,

46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69,  
70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160,  
170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 350, 400, 450, 500,  
5 550, 600, 650, 700, 750, 800, 850, 900, 950, or 1000 contiguous nucleotides in length  
from the polynucleotides of (A) – (E). Optionally, the number of such subsequences  
encoded by a polynucleotide of the instant embodiment can be any integer selected from  
the group consisting of from 1 to 1000, such as 2, 3, 4, or 5. The subsequences can be  
separated by any integer of nucleotides from 1 to the number of nucleotides in the  
10 sequence such as at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 200, 300, 400, 500,  
600, 700, 800, 900, or 1000 nucleotides.

The subsequences of the present invention can comprise structural characteristics  
of the sequence from which it is derived. Alternatively, the subsequences can lack  
certain structural characteristics of the larger sequence from which it is derived such as  
15 poly (A) tail. Optionally, a subsequence from a polynucleotide encoding a polypeptide  
having at least one linear epitope in common with a prototype polypeptide sequence as  
provided in SEQ ID NO:2, may encode an epitope in common with the prototype  
sequence. Alternatively, the subsequence may not encode an epitope in common with the  
prototype sequence but can be used to isolate the larger sequence by, for example, nucleic  
20 acid hybridization with the sequence from which it's derived. Subsequences can be used  
to modulate or detect gene expression by introducing into the subsequence compounds,  
which bind, intercalate, cleave and/or crosslink to nucleic acids. Exemplary compounds  
include acridine, psoralen, phenanthroline, naphthoquinone, daunomycin, or  
chloroethylaminoaryl conjugates.

25 The subsequences or fragments of the polynucleotides of the present invention  
may encode protein fragments that retain the biological activity of the native protein and  
hence have PRI-C10-like activity. Examples of PRI-C10-like activity include  
participation in the pathogenic response of the plant and enhancement of disease  
resistance. Alternatively, fragments of a nucleotide sequence that are useful as

hybridization probes generally do not encode fragment proteins retaining biological activity. A fragment of a PR1-C10 nucleotide sequence that encodes a biologically active portion of a PR1 protein of the invention will encode at least 15, 25, 30, 50, 100, 150, 200, or up to 203 contiguous amino acids of the full-length PR1-C10 protein of the invention. Fragments of a PR1-C10 nucleotide sequence that are useful as hybridization probes or PCR primers generally need not encode a biologically active portion of a PR1-C10 protein.

Thus, a fragment of a PR1-C10 nucleotide sequence may encode a biologically active portion of a PR1-C10 protein, or it may be a fragment that can be used as a hybridization probe or PCR primer using methods disclosed below. A biologically active portion of a PR1-C10 protein can be prepared by isolating a portion of one of the PR1-C10 nucleotide sequences of the invention, expressing the encoded portion of the PR1-C10 protein (e.g., by recombinant expression *in vitro*), and assessing the activity of the encoded portion of the PR1-C10 protein. Nucleic acid molecules that are fragments of a PR1-C10 nucleotide sequence comprise at least 16, 20, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 800, or up to 898 nucleotides present in full length PR1-C10 disclosed in SEQ ID NO:1 or at least 16, 20, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, or up to 612 nucleotides present in SEQ ID NO:3, (nucleotides 63-674 of SEQ ID NO:1).

By “variants” is intended substantially similar sequences. For nucleotide sequences, conservative variants include those sequences that, because of the degeneracy of the genetic code, encode the amino acid sequence of one of the PR1-C10 polypeptides of the invention. Naturally occurring allelic variants such as these can be identified with the use of well-known molecular biology techniques, as, for example, with polymerase chain reaction (PCR) and hybridization techniques as outlined below. Variant nucleotide sequences also include synthetically derived nucleotide sequences, such as those generated, for example, by using site-directed mutagenesis but which still encode a PR1-C10 protein of the invention. Generally, variants of a particular nucleotide sequence of

the invention will have at least about 60%, 65%, 70%, generally at least about 75%, 80%, 85%, preferably at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, and more preferably at least about 98%, 99% or more sequence identity to that particular nucleotide sequence as determined by sequence alignment programs described elsewhere herein  
5 using default parameters.

### Construction of Nucleic Acids

The isolated nucleic acids of the present invention can be made using (a) standard recombinant methods, (b) synthetic techniques, or combinations thereof. In some  
10 embodiments, the polynucleotides of the present invention will be cloned, amplified, or otherwise constructed from a monocot. In one embodiment the monocot is *Zea mays*.

The nucleic acids may conveniently comprise sequences in addition to a polynucleotide of the present invention. For example, a multi-cloning site comprising one or more endonuclease restriction sites may be inserted into the nucleic acid to aid in  
15 isolation of the polynucleotide. Also, translatable sequences may be inserted to aid in the isolation of the translated polynucleotide of the present invention. For example, a hexahistidine marker sequence provides a convenient means to purify the proteins of the present invention. A polynucleotide of the present invention can be attached to a vector, adapter, or linker for cloning and/or expression of a polynucleotide of the present  
20 invention. Additional sequences may be added to such cloning and/or expression sequences to optimize their function in cloning and/or expression, to aid in isolation of the polynucleotide, or to improve the introduction of the polynucleotide into a cell. Typically, the length of a nucleic acid of the present invention is less than 20 kilobase pairs, often less than 15 kb, and frequently less than 10 kb. Use of cloning vectors,  
25 expression vectors, adapters, and linkers is well known and extensively described in the art. For a description of various nucleic acids see, for example, Stratagene Cloning Systems, Catalogs 1995, 1996, 1997 (La Jolla, CA); and, Amersham Life Sciences, Inc, Catalog 1997 (Arlington Heights, IL).

### A. Recombinant Methods for Constructing Nucleic Acids

The isolated nucleic acid compositions of this invention, such as RNA, cDNA, genomic DNA, or a hybrid thereof, can be obtained from plant biological sources using any number of cloning methodologies known to those of skill in the art. In some  
5   embodiments, oligonucleotide probes that selectively hybridize, under stringent conditions, to the polynucleotides of the present invention are used to identify the desired sequence in a cDNA or genomic DNA library. While isolation of RNA, and construction of cDNA and genomic libraries is well known to those of ordinary skill in the art, the following highlights some of the methods employed.

#### 10   A1. mRNA Isolation and Purification

Total RNA from plant cells comprises such nucleic acids as mitochondrial RNA, chloroplastic RNA, rRNA, tRNA, hnRNA and mRNA. Total RNA preparation typically involves lysis of cells and removal of organelles and proteins, followed by precipitation  
15   of nucleic acids. Extraction of total RNA from plant cells can be accomplished by a variety of means. Frequently, extraction buffers include a strong detergent such as SDS and an organic denaturant such as guanidinium isothiocyanate, guanidine hydrochloride or phenol. Following total RNA isolation, poly(A)<sup>+</sup> mRNA is typically purified from the remainder RNA using oligo(dT) cellulose. Exemplary total RNA and mRNA isolation  
20   protocols are described in Clark *et al.* (1997) *Plant Molecular Biology: A Laboratory Manual*, Springer-Verlag, Berlin; and, Ausubel, *et al.* (1995) *Current Protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York. Total RNA and mRNA isolation kits are commercially available from vendors such as Stratagene (La Jolla, CA), Clontech (Palo Alto, CA), Pharmacia (Piscataway, NJ), and 5'-3' (Paoli Inc.,  
25   PA). See also, U.S. Patent Nos. 5,614,391; and, 5,459,253. The mRNA can be fractionated into populations with size ranges of about 0.5, 1.0, 1.5, 2.0, 2.5, or 3.0 kb. The cDNA synthesized for each of these fractions can be size selected to the same size range as its mRNA prior to vector insertion. This method helps eliminate truncated cDNA formed by incompletely reverse transcribed mRNA.

## A2. Construction of a cDNA Library

Construction of a cDNA library generally entails five steps. First, first strand cDNA synthesis is initiated from a poly(A)+ mRNA template using a poly(dT) primer or random hexanucleotides. Second, the resultant RNA-DNA hybrid is converted into  
5 double stranded cDNA, typically by reaction with a combination of RNase H and DNA polymerase I (or Klenow fragment). Third, the termini of the double stranded cDNA are ligated to adaptors. Ligation of the adaptors can produce cohesive ends for cloning. Fourth, size selection of the double stranded cDNA eliminates excess adaptors and primer fragments, and eliminates partial cDNA molecules due to degradation of mRNAs  
10 or the failure of reverse transcriptase to synthesize complete first strands. Fifth, the cDNAs are ligated into cloning vectors and packaged. cDNA synthesis protocols are well known to the skilled artisan and are described in such standard references as: Clark *et al.* (1997) *Plant Molecular Biology: A Laboratory Manual*, Springer-Verlag, Berlin; and, Ausubel, *et al.* (1995) *Current Protocols in molecular Biology*, Greene Publishing  
15 and Wiley-Interscience, New York. cDNA synthesis kits are available from a variety of commercial vendors such as Stratagene or Pharmacia.

A number of cDNA synthesis protocols have been described which provide substantially pure full-length cDNA libraries. Substantially pure full-length cDNA libraries are constructed to comprise at least 90%, and more preferably at least 93% or  
20 95% full-length inserts amongst clones containing inserts. The length of insert in such libraries can be from 0 to 8, 9, 10, 11, 12, 13, or more kilobase pairs. Vectors to accommodate inserts of these sizes are known in the art and available commercially. See, e.g., Stratagene's lambda ZAP Express (cDNA cloning vector with 0 to 12 kb cloning capacity).

25 An exemplary method of constructing a greater than 95% pure full-length cDNA library is described by Carninci *et al.* (1996) *Genomics*, 37:327-336. In that protocol, the cap-structure of eukaryotic mRNA is chemically labeled with biotin. By using streptavidin-coated magnetic beads, only the full-length first-strand cDNA/mRNA hybrids are selectively recovered after RNase I treatment. The method provides a high

yield library with an unbiased representation of the starting mRNA population. Other methods for producing full-length libraries are known in the art. See, e.g., Edery *et al.* (1995) *Mol. Cell Biol.* 15(6):3363-3371 and PCT Application WO 96/34981.

5     A3. Normalized or Subtracted cDNA Libraries

A non-normalized cDNA library represents the mRNA population of the tissue it was made from. Since unique clones are out-numbered by clones derived from highly expressed genes their isolation can be laborious. Normalization of a cDNA library is the process of creating a library in which each clone is more equally represented.

10     A number of approaches to normalize cDNA libraries are known in the art. One approach is based on hybridization to genomic DNA. The frequency of each hybridized cDNA in the resulting normalized library would be proportional to that of each corresponding gene in the genomic DNA. Another approach is based on kinetics. If cDNA reannealing follows second-order kinetics, rarer species anneal less rapidly and the  
15     remaining single-stranded fraction of cDNA becomes progressively more normalized during the course of the hybridization. Specific loss of any species of cDNA, regardless of its abundance, does not occur at any Cot value. Construction of normalized libraries is described in Ko (1990) *Nucl. Acids. Res.*, 18(19):5705-5711; Patanjali *et al.* (1991) *Proc. Natl. Acad. U.S.A.*, 88:1943-1947; U.S. Patents 5,482,685; and 5,637,685. In an  
20     exemplary method described by Soares *et al.*, normalization resulted in reduction of the abundance of clones from a range of four orders of magnitude to a narrow range of only 1 order of magnitude (Soares *et al.* (1994) *Proc. Natl. Acad. Sci. USA*, 91:9228-9232).

Subtracted cDNA libraries are another means to increase the proportion of less abundant cDNA species. In this procedure, cDNA prepared from one pool of mRNA is  
25     depleted of sequences present in a second pool of mRNA by hybridization. The cDNA:mRNA hybrids are removed and the remaining un-hybridized cDNA pool is enriched for sequences unique to that pool. See, Foote *et al.* (1997) *Plant Molecular Biology: A Laboratory Manual*, Springer-Verlag, Berlin; Kho and Zarbl (1991) *Technique*, 3(2):58-63; Sive and St. John (1998) *Nucl. Acids Res.*, 16(22):10937;



Ausubel, *et al.* (1995) *Current Protocols in Molecular Biology*, Greene publishing and Wiley-Interscience, New York; and, Swaroop *et al.*, (1991) *Nucl. Acids Res.* 19(8):1954. cDNA subtraction kits are commercially available. See e.g., PCR-Select (Clontech, Palo Alto, CA).

5

#### A4. Construction of a Genomic Library

To construct genomic libraries, large segments of genomic DNA are generated by fragmentation, e.g. using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. Methodologies to accomplish these ends, and sequencing methods to verify the sequence of nucleic acids are well known in the art. Examples of appropriate molecular biological techniques and instructions sufficient to direct persons of skill through many construction, cloning, and screening methodologies are found in Sambrook *et al.* (1989) *Molecular cloning: A Laboratory Manual*, 2<sup>nd</sup> Ed., Cold Spring Harbor Laboratory Vols. 1-3; Berger, *et al.* (1987) *Methods in Enzymology*, Vol. 152: *Guide to Molecular Cloning Techniques*, San Diego: Academy Press, Inc.; Ausubel, *et al.* (1995) *Current protocols in Molecular Biology*, Greene Publishing and Wiley-Interscience, New York; Clark, *et al.* (1997) *Plant Molecular Biology: A Laboratory Manual*, Springer-Verlag, Berlin. Kits for construction of genomic libraries are also commercially available.

10

15

20

#### A5. Nucleic Acid Screening and Isolation Methods

The cDNA or genomic library can be screened using a probe based upon the sequence of a polynucleotide of the present invention such as those disclosed herein. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. Those of skill in the art will appreciate that various degrees of stringency of hybridization can be employed in the assay; and the hybridization of the wash medium can be stringent or mild. As the conditions for hybridization become more stringent, there must be a greater degree of complementarity between the probe and the target for duplex formation to occur. The

25

degree of stringency can be controlled by temperature, ionic strength, pH and the presence of a partially denaturing solvent such as formamide. For example, changing the polarity of the reactant solution through manipulation of the concentration of formamide within the range of 0% to 50% conveniently varies the stringency of hybridization. The degree of complementarity (sequence identity) required for detectable binding will vary in accordance with the stringency of the hybridization medium and/ or wash medium. The degree of complementarity will optimally be 100 percent, however, it should be understood that reducing the stringency of the hybridization and/or wash medium might compensate for minor sequence variations in the probes and primers.

The nucleic acids of interest can also be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of polynucleotides of the present invention and related genes directly from genomic DNA or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes. Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods are found in Berger, Sambrook, and Ausubel, as well as Mullis *et al.*, U.S. Patent No. 4,683,202 (1987); and, Innis, *et al.* (1990) *PCR Protocols A guide to Methods and Applications*, Academic press Inc., San Diego, CA (1990). Commercially available kits for genomic PCR amplification are known in the art. See, e.g., Advantage-GC Genomic PCR Kit (Clontech). The T4 gene 32 protein (Boehringer Mannheim) can be used to improve yield of long PCR products.

PCR-based screening methods have also been described. Wilfinger *et al.* (1997) *BioTechniques* 22(3):481-486 describes a PCR-based method in which the longest cDNA is identified in the first step so those incomplete clones can be eliminated from study. In that method, a primer pair is synthesized with one primer annealing to the 5' end of the sense strand of the desired cDNA and the other primer to the vector. Clones are pooled to allow large-scale screening. By this procedure, the longest possible clone is identified

amongst candidate clones. Further, the PCR product is used solely as a diagnostic for the presence of the desired cDNA and does not utilize the PCR product itself. Such methods are particularly effective in combination with a full-length cDNA construction methodology described above.

5

#### *B. Synthetic Methods for Constructing Nucleic Acids*

The isolated nucleic acids of the present invention can also be prepared by direct chemical synthesis by methods such as the phosphotriester method of Narang *et al.* (1979) *Meth. Enzymol.* 68: 90-99; the phosphodiester method of Brown *et al.* (1979) *Meth. Enzymol.* 68:109-151; the diethylphosphoramidite method of Beaucage *et al.* (1981) *Tetra. Lett.* 22: 1859-1862; the solid phase phosphoramidite triester method described by Beaucage and Caruthers (1981) *Tetra. Letts.* 22(20): 1859-1862, e.g., using an automated synthesizer, e.g., as described in Needham-VanDevanter *et al.* (1984) *Nucleic Acids Res.* 12:6159-6168; and, the solid support method of U.S. Patent No.

10

15

4,458,066. Chemical synthesis generally produces a single stranded oligonucleotide. This may be converted into double stranded DNA by hybridization with a complementary sequence, or by polymerization with a DNA polymerase using the single strand as a template. One of skill will recognize that while chemical synthesis of DNA is best employed for sequences of about 100 bases or less, longer sequences may be obtained by the ligation of shorter sequences.

20

#### Recombinant Expression Cassettes

The present invention further provides recombinant expression cassettes comprising a nucleic acid of the present invention. A nucleic acid sequence coding for the desired polynucleotide of the present invention or a fragment or variant thereof, for example a cDNA or a genomic sequence encoding a full length polypeptide of the present invention, can be used to construct a recombinant expression cassette which can be introduced into the desired host cell. A recombinant expression cassette will typically comprise a polynucleotide of the present invention operably linked to transcriptional

25

initiation regulatory sequences which will direct the transcription of the polynucleotide in the intended host cell, such as tissues of a transformed plant.

Plant expression vectors may include (1) a cloned plant gene under the transcriptional control of 5' and 3' regulatory sequences and (2) a dominant selectable  
5 marker. Such plant expression vectors may also contain, if desired, a promoter regulatory region (e.g., one conferring inducible or constitutive, environmentally- or developmentally-regulated, or cell- or tissue-preferred/selective expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

10 For example, the PR1-C10 sequences of the invention are provided in expression cassettes for expression in the plant of interest. The cassette will include 5' and 3' regulatory sequences operably linked to a PR1-C10 sequence of the invention. By “operably linked” is intended a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates and mediates transcription of the DNA  
15 sequence corresponding to the second sequence. Generally, operably linked means that the nucleic acid sequences being linked are contiguous and, where necessary to join two protein coding regions, contiguous and in the same reading frame. The cassette may additionally contain at least one additional gene to be cotransformed into the organism. Alternatively, the additional gene(s) can be provided on multiple expression cassettes.

20 Such an expression cassette is provided with a plurality of restriction sites for insertion of the PR1-C10 sequence to be under the transcriptional regulation of the regulatory regions. The expression cassette may additionally contain selectable marker genes.

The expression cassette will include in the 5'-3' direction of transcription, a  
25 transcriptional and translational initiation region, a PR1-C10 DNA sequence of the invention, and a transcriptional and translational termination region functional in plants. The transcriptional initiation region, the promoter, may be native or analogous or foreign or heterologous to the plant host. Additionally, the promoter may be the natural sequence or alternatively a synthetic sequence. By “foreign” is intended that the transcriptional

initiation region is not found in the native plant into which the transcriptional initiation region is introduced. As used herein, a chimeric gene comprises a coding sequence operably linked to a transcription initiation region that is heterologous to the coding sequence.

5           While it may be preferable to express the sequences using heterologous promoters, the native promoter sequences may be used. Such constructs would change expression levels of PR1-C10 in the plant or plant cell. Thus, the phenotype of the plant or plant cell is altered.

10           The termination region may be native with the transcriptional initiation region, may be native with the operably linked DNA sequence of interest, or may be derived from another source. Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also Guerineau *et al.* (1991) *Mol. Gen. Genet.* 262:141-144; Proudfoot (1991) *Cell* 64:671-674; Sanfacon *et al.* (1991) *Genes Dev.* 5:141-149; Mogen *et al.* (1990) *Plant Cell* 2:1261-1272; Munroe *et al.* (1990) *Gene* 91:151-158; Ballas *et al.* (1989) *Nucleic Acids Res.* 17:7891-7903; and Joshi *et al.* (1987) *Nucleic Acid Res.* 15:9627-9639.

20           Where appropriate, the gene(s) may be optimized for increased expression in the transformed plant. That is, the genes can be synthesized using plant-preferred codons for improved expression. See, for example, Campbell and Gowri (1990) *Plant Physiol.* 92:1-11 for a discussion of host-preferred codon usage. Methods are available in the art for synthesizing plant-preferred genes. See, for example, U.S. Patent Nos. 5,380,831, and 5,436,391, and Murray *et al.* (1989) *Nucleic Acids Res.* 17:477-498, herein incorporated by reference.

25           Additional sequence modifications are known to enhance gene expression in a cellular host. These include elimination of sequences encoding spurious polyadenylation signals, exon-intron splice site signals, transposon-like repeats, and other such well-characterized sequences that may be deleterious to gene expression. The G-C content of the sequence may be adjusted to levels average for a given cellular host, as calculated by

reference to known genes expressed in the host cell. When possible, the sequence is modified to avoid predicted hairpin secondary mRNA structures.

The expression cassettes may additionally contain 5' leader sequences in the expression cassette construct. Such leader sequences can act to enhance translation.

- 5 Translation leaders are known in the art and include: picornavirus leaders, for example, EMCV leader (Encephalomyocarditis 5' noncoding region) (Elroy-Stein *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6126-6130); potyvirus leaders, for example, TEV leader (Tobacco Etch Virus) (Gallie *et al.* (1995) *Gene* 165(2):233-238), MDMV leader (Maize Dwarf Mosaic Virus) (*Virology* 154:9-20), and human immunoglobulin heavy-chain
- 10 binding protein (BiP) (Macejak *et al.* (1991) *Nature* 353:90-94); untranslated leader from the coat protein mRNA of alfalfa mosaic virus (AMV RNA 4) (Jobling *et al.* (1987) *Nature* 325:622-625); tobacco mosaic virus leader (TMV) (Gallie *et al.* (1989) in *Molecular Biology of RNA*, ed. Cech (Liss, New York), pp. 237-256); and maize chlorotic mottle virus leader (MCMV) (Lommel *et al.* (1991) *Virology* 81:382-385). See
- 15 also, Della-Cioppa *et al.* (1987) *Plant Physiol.* 84:965-968. Other methods known to enhance translation can also be utilized, for example, introns, and the like.

- In preparing the expression cassette, the various DNA fragments may be manipulated, so as to provide for the DNA sequences in the proper orientation and, as appropriate, in the proper reading frame. Toward this end, adapters or linkers may be
- 20 employed to join the DNA fragments or other manipulations may be involved to provide for convenient restriction sites, removal of superfluous DNA, removal of restriction sites, or the like. For this purpose, *in vitro* mutagenesis, primer repair, restriction, annealing, resubstitutions, e.g., transitions and transversions, may be involved.

- A number of promoters can be used in the practice of the invention. A number of
- 25 promoters can be used in the practice of the invention. The promoters can be selected based on the desired outcome. The nucleic acids can be combined with constitutive, tissue-preferred, or other promoters for expression in plants. A plant promoter fragment can be employed which will direct expression of a polynucleotide of the present invention in all tissues of a regenerated plant. Such promoters are referred to herein as

"constitutive" promoters and are active under most environmental conditions and stated of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region (Odell *et al.* (1985) *Nature* 313:810-812), the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, the ubiquitin 1 promoter (Christensen *et al.* (1992) *Plant Mol Biol* 18, 675-689; Bruce *et al.* (1989) *Proc Natl Acad Sci USA* 86, 9692-9696), the Smas promoter, the cinnamyl alcohol dehydrogenase promoter (U.S. Patent No, 5,683,439), the *Nos* promoter, the pEmu promoter (Last *et al.* (1991) *Theor. Appl. Genet.* 81:581-588), the rubisco promoter, the GRP1-8 promoter, the maize constitutive promoters described in PCT Publication No. WO 99/43797 which include the histone H2B, metallothionein, alpha-tubulin 3, elongation factor efla, ribosomal protein rps8, chlorophyll a/b binding protein, and glyceraldehyde-3-phosphate dehydrogenase promoters, the core promoter of the Rsyn7 promoter and other constitutive promoters disclosed in WO 99/43838 and U.S. Patent No. 6,072,050; rice actin (McElroy *et al.* (1990) *Plant Cell* 2:163-171); ubiquitin (Christensen *et al.* (1989) *Plant Mol. Biol.* 12:619-632 and Christensen *et al.* (1992) *Plant Mol. Biol.* 18:675-689); MAS (Velten *et al.* (1984) *EMBO J.* 3:2723-2730); ALS promoter (U.S. Patent No. 5,659,026), and the like. Other constitutive promoters include, for example, U.S. Patent Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; and 5,608,142. For constitutive expression of the polynucleotide of the present invention, the ubiquitin 1 promoter is the preferred promoter.

Where low level expression is desired, weak promoters will be used. It is recognized that weak inducible promoters may be used. Additionally, either a weak constitutive or a weak tissue specific promoter may be used. Generally, by a "weak promoter" is intended a promoter that drives expression of a coding sequence at a low level. By low level is intended at levels of about 1/1000 transcripts to about 1/100,000 transcripts to about 1/500,000 transcripts. Alternatively, it is recognized that weak promoters also encompass promoters that are expressed in only a few cells and not in others to give a total low level of expression. Such weak constitutive promoters include, for example, the core promoter of the Rsyn7 (PCT Publication No. WO 97/44756, WO

99/43838 and U.S. Patent No. 6,072,050), core 35S CaMV promoter, and the like. Other constitutive promoters include, for example, U.S. Patent Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; and 5,608,142. For constitutive expression of the polynucleotide of the present invention, the ubiquitin 1 promoter is the preferred promoter. See also, U.S. Patent No. 6,177,611, and herein incorporated by reference.

Where a promoter is expressed at unacceptably high levels, portions of the promoter sequence can be deleted or modified to decrease expression levels. Additionally, to obtain a varied series in the level of expression, one can also make a set of transgenic plants containing the polynucleotides of the present invention with a strong constitutive promoter, and then rank the transgenic plants according to the observed level of expression. The transgenic plants will show a variety in performance, from high expression to low expression. Factors such as chromosomal position effect, cosuppression, and the like will affect the expression of the polynucleotide.

Alternatively, the plant promoter can direct expression of a polynucleotide of the present invention under environmental control. Such promoters are referred to here as "inducible" promoters. Environmental conditions that may effect transcription by inducible promoters include pathogen attack, anaerobic conditions, or the presence of light. Examples of inducible promoters are the Adh1 promoter, which is inducible by hypoxia or cold stress, the Hsp70 promoter, which is inducible by heat stress, and the PPK promoter, which is inducible by light. Examples of pathogen-inducible promoters include those from proteins, which are induced following infection by a pathogen; *e.g.*, pathogen-related proteins (PR proteins), SAR proteins, beta-a, 3-glucanase, chitinase, etc. See, for example, Redolfi, *et al.* (1983) *Meth J. Plant Pathol.* 89:245-254; Uknes *et al.* (1992) *The Plant Cell* 4:645-656; Van Loon (1985) *Plant Mol. Virol.* 4:111-116; and PCT Publication No. WO 99/43819. See also the copending application entitled "Inducible Maize Promoters," WO 99/43819, published Sept. 9, 1999, herein incorporated by reference.



Of interest are promoters that are expressed locally at or near the site of pathogen infection. See, for example, Marineau *et al.* (1987) *Plant Mol. Biol.* 9:335-342; Matton *et al.* (1989) *Molecular Plant-Microbe Interactions* 2:325-331; Somsisch *et al.* (1986) *Proc. Natl. Acad. Sci. USA* 83:2427-2430; Somsisch *et al.* (1988) *Mol. Gen. Genet.* 2:93-98; and Yang (1996) *Proc. Natl. Acad. Sci. USA* 93:14972-14977. See also, Chen *et al.* (1996) *Plant J.* 10:955-966; Zhang *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:2507-2511; Warner *et al.* (1993) *Plant J.* 3:191-201; Siebertz *et al.* (1989) *Plant Cell* 1:961-968; U.S. Patent No. 5,750,386 (nematode-inducible); and the references cited therein. Of particular interest is the inducible promoter for the maize PRms gene, whose expression is induced by the pathogen *Fusarium moniliforme* (see, for example, Cordero *et al.* (1992) *Physiol. Mol. Plant Path.* 41:189-200).

Additionally, as pathogens find entry into plants through wounds or insect damage, a wound inducible promoter may be used in the constructs of the invention. Such wound inducible promoter include potato proteinase inhibitor (pin II) gene (Ryan (1990) *Annu Rev Phytopath* 28:425-449; Duan *et al.* (1996) *Nat Biotech* 14:494-498); wun1 and wun 2, US Patent No. 5,428,148; win1 and win2 (Stanford *et al.* (1989) *Mol Gen Genet* 215:200-208); systemin (McGurl *et al.* (1992) *Science* 225:1570-1573); WIP1 (Rohmeier *et al.* (1993) *Plant Mol Biol* 22:783-792; Eckelkamp *et al.* (1993) *FEB Letters* 323:73-76); MPI gene (Cordero *et al.* (1994) *The Plant J* 6(2):141-150); and the like, herein incorporated by reference.

Examples of promoters under developmental control include promoters that initiate transcription only, or preferentially, in certain tissues, such as leaves, roots, fruit, seeds, or flowers. Exemplary promoters include the anther specific promoter 5126 (U.S. Patent Nos. 5,689,049 and 5,689,051), glob-1 promoter, and gamma-zein promoter. An exemplary promoter for leaf- and stalk-preferred expression is MS8-15 (WO 98/00533). Examples of seed-preferred promoters included, but are not limited to, 27 kD gamma zein promoter and waxy promoter (Boronat *et al.* (1986) *Plant Sci* 47:95-102; Reina *et al.* (1990) *Nucleic Acids Res* 18(21):6426; and Kloesgen *et al.* (1986) *Mol Gen Genet* 203:237-244). Promoters that express in the embryo, pericarp, and endosperm are

disclosed in PCT publication WO 00/11177 published March 2, 2000 and WO 00/12733, published March 9, 2000, both of which are hereby incorporated by reference. The operation of a promoter may also vary depending on its location in the genome. Other tissue preferred promoters include Yamamoto *et al.* (1997) *Plant J.* 12(2):255-265;

- 5 Kawamata *et al.* (1997) *Plant Cell Physiol.* 38(7):792-803; Hansen *et al.* (1997) *Mol. Gen Genet.* 254(3):337-343; Russell *et al.* (1997) *Transgenic Res.* 6(2):157-168; Rinehart *et al.* (1996) *Plant Physiol.* 112(3):1331-1341; Van Camp *et al.* (1996) *Plant Physiol.* 112(2):525-535; Canevascini *et al.* (1996) *Plant Physiol.* 112(2):513-524; Yamamoto *et al.* (1994) *Plant Cell Physiol.* 35(5):773-778; Lam (1994) *Results Probl. Cell Differ.* 20:181-196; Orozco *et al.* (1993) *Plant Mol Biol.* 23(6):1129-1138; Matsuoka *et al.* (1993) *Proc Natl. Acad. Sci. USA* 90(20):9586-9590; and Guevara-Garcia *et al.* (1993) *Plant J.* 4(3):495-505. Such promoters can be modified, if necessary, for weak expression. Thus, a developmentally regulated promoter may become fully or partially constitutive in certain locations. A developmentally regulated promoter can also be
- 10
- 15 modified, if necessary, for weak expression.

Both heterologous and non-heterologous (i.e. endogenous) promoters can be employed to direct expression of the nucleic acids of the present invention. These promoters can also be used, for example, in recombinant expression cassettes to drive expression of antisense nucleic acids to reduce, increase, or alter concentration and/or composition of the proteins of the present invention in a desired tissue. Thus, in some

20

embodiments, the nucleic acid construct will comprise a promoter functional in a plant cell, such as in *Zea mays* operably linked to a polynucleotide of the present invention. Promoters useful in these embodiments include the endogenous promoters driving expression of a polypeptide of the present invention.

- 25 In some embodiments, isolated nucleic acids which serve as a promoter or enhancer elements can be introduced in the appropriate position (generally upstream) of a non-heterologous form of a polynucleotide of the present invention so as to up or down regulate expression of a polynucleotide of the present invention. For example, endogenous promoters can be altered *in vivo* by mutation, deletion, and/or substitution

(see, Kmiec, U.S. Patent 5,565,350; Zarling *et al.*, PCT/US93/03868), or isolated promoters can be introduced into a plant cell in the proper orientation and distance from a gene of the present invention so as to control the expression of the gene. Gene expression can be modulated under conditions suitable for plant growth so as to alter the total concentration and/or alter the composition of the polypeptides of the present invention in plant cell. Thus, the present invention provides compositions, and methods for making, heterologous promoters and/or enhancers operably linked to a native, endogenous (i.e., non-heterologous) form of a polynucleotide of the present invention.

If polypeptide expression is desired, it is generally desirable to include a polyadenylation region at the 3'-end of a polynucleotide coding region. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA. The 3' end sequence to be added can be derived from, for example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or less preferably from any other eukaryotic gene.

An intron sequence can be added to the 5' untranslated region or the coding sequence of the partial coding sequence to increase the amount of the mature message that accumulates in the cytosol. Inclusion of a spliceable intron in the transcription unit in both plant and animal expression constructs has been shown to increase gene expression at both the mRNA and protein levels up to 1000-fold, Buchman and Berg (1988) *Mol. Cell biol.* 8: 4395-4405; Callis *et al.* (1987) *Genes Dev.* 1: 1183-1200. Such intron enhancement of gene expression is typically greatest when placed near the 5' end of the transcription unit. Use of the maize introns Adh1-S intron 1, 2, and 6, and the Bronze-1 intron are known in the art. See generally, Freeling *et al.* (1994) *The Maize Handbook*, Chapter 116, Springer, New York.

The vector comprising the sequences from a polynucleotide of the present invention will typically comprise a marker gene, which confers a selectable phenotype on plant cells. Usually, the selectable marker gene will encode antibiotic resistance, with suitable genes including genes coding for resistance to the antibiotic spectinomycin (e.g., the *aada* gene), the streptomycin phosphotransferase (SPT) gene coding for streptomycin

resistance, the neomycin phosphotransferase (NPTII) gene encoding kanamycin or geneticin resistance, the hygromycin phosphotransferase (HPT) gene coding for hygromycin resistance, genes coding for resistance to herbicides which act to inhibit the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides (e.g., the acetolactate synthase (ALS) gene containing mutations leading to such resistance in particular the S4 and/or Hra mutations), genes coding for resistance to herbicides which act to inhibit action of glutamine synthase, such as phosphinothricin or basta (e.g., the *bar* gene), or other such genes known in the art. The *bar* gene encodes resistance to the herbicide basta, the *nptII* gene encodes resistance to the antibiotics kanamycin and geneticin, and the ALS gene encodes resistance to the herbicide chlorsulfuron.

Generally, the expression cassette will comprise a selectable marker gene for the selection of transformed cells. Selectable marker genes are utilized for the selection of transformed cells or tissues. Marker genes include genes encoding antibiotic resistance, such as those encoding neomycin phosphotransferase II (NEO) and hygromycin phosphotransferase (HPT), as well as genes conferring resistance to herbicidal compounds, such as glufosinate ammonium, bromoxynil, imidazolinones, and 2,4-dichlorophenoxyacetate (2,4-D). See generally, Yarranton (1992) *Curr. Opin. Biotech.* 3:506-511; Christopherson *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6314-6318; Yao *et al.* (1992) *Cell* 71:63-72; Reznikoff (1992) *Mol. Microbiol.* 6:2419-2422; Barkley *et al.* (1980) in *The Operon*, pp. 177-220; Hu *et al.* (1987) *Cell* 48:555-566; Brown *et al.* (1987) *Cell* 49:603-612; Figge *et al.* (1988) *Cell* 52:713-722; Deuschle *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:5400-5404; Fuerst *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:2549-2553; Deuschle *et al.* (1990) *Science* 248:480-483; Gossen (1993) Ph.D. Thesis, University of Heidelberg; Reines *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:1917-1921; Labow *et al.* (1990) *Mol. Cell. Biol.* 10:3343-3356; Zambretti *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:3952-3956; Baim *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:5072-5076; Wyborski *et al.* (1991) *Nucleic Acids Res.* 19:4647-4653; Hillenand-Wissman (1989) *Topics Mol. Struc. Biol.* 10:143-162; Degenkolb *et al.* (1991) *Antimicrob. Agents Chemother.* 35:1591-1595; Kleinschmidt *et al.* (1988) *Biochemistry* 27:1094-1104; Bonin (1993) Ph.D. Thesis,

University of Heidelberg; Gossen *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Oliva *et al.* (1992) *Antimicrob. Agents Chemother.* 36:913-919; Hlavka *et al.* (1985) *Handbook of Experimental Pharmacology*, Vol. 78 (Springer-Verlag, Berlin); Gill *et al.* (1988) *Nature* 334:721-724. Such disclosures are herein incorporated by reference.

5           The above list of selectable marker genes is not meant to be limiting. Any selectable marker gene can be used in the present invention.

Typical vectors useful for expression of genes in higher plants are well known in the art and include vectors derived from the tumor-induced (Ti) plasmid of *Agrobacterium tumefaciens* described by Rogers *et al.* (1987) *Meth. In Enzymol.*

10   153:253-277. These vectors are plant integrating vectors in that upon transformation, the vectors integrate a portion of vector DNA into the genome of the host plant. Exemplary *A. tumefaciens* vectors useful herein are plasmids pKYLX6 and pKYLX7 of Schardl *et al.* (1987) *Gene*, 61:1-11 and Berger *et al.* (1989) *Proc. Natl. Acad. Sci. U.S.A.*, 86:8402-8406. Another useful vector herein is plasmid pBI101.2 that is available from Clontech  
15   Laboratories, Inc. (Palo Alto, CA).

A polynucleotide of the present invention can be expressed in either sense or anti-sense orientation as desired. It will be appreciated that control of gene expression in either sense or antisense orientation can have a direct impact on the observable plant characteristics. Antisense technology can be conveniently used to inhibit gene expression  
20   in plants. To accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the antisense strand of RNA will be transcribed. The construct is then transformed into plants and the antisense strand of RNA is produced. In plant cells, it has been shown that antisense RNA inhibits gene expression by preventing the accumulation of mRNA which encodes the enzyme of interest, see,  
25   e.g., Sheehy *et al.* (1988) *Proc. Nat'l. Acad. Sci (USA)* 85:8805-8809; and Hiatt *et al.* U.S. Patent No. 4,801,340.

Antisense nucleotides are constructed to hybridize with the corresponding mRNA. Modifications of the antisense sequences may be made as long as the sequences hybridize to and interfere with expression of the corresponding mRNA. In this manner, antisense

5 constructions having 70%, preferably 80%, more preferably 85% sequence identity to the corresponding antisensed sequences may be used. Furthermore, portions of the antisense nucleotides may be used to disrupt the expression of the target gene. Generally, sequences of at least 50 nucleotides, 100 nucleotides, 200 nucleotides, or greater may be used.

Another method of suppression is sense suppression. Introduction of nucleic acid configured in the sense orientation has been shown to be an effective means by which to block the transcription of target genes. For an example of the use of this method to modulate expression of endogenous genes see, Napoli *et al.* (1990) *The Plant Cell* 2:279-289 and U.S. Patent No. 5,034,323. The methods generally involve transforming plants with a DNA construct comprising a promoter that drives expression in a plant operably linked to at least a portion of a nucleotide sequence that corresponds to the transcript of the endogenous gene. Typically, such a nucleotide sequence has substantial sequence identity to the sequence of the transcript of the endogenous gene, preferably greater than about 65% sequence identity, more preferably greater than about 85% sequence identity, most preferably greater than about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity. See, U.S. Patent Nos. 5,283,184 and 5,034,323; herein incorporated by reference.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of plant genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs. The design and use of target RNA-specific ribozymes is described in Haseloff *et al.* (1988) *Nature* 334:585-591.

A variety of cross-linking agents, alkylating agents and radical generating species as pendant groups on polynucleotides of the present invention can be used to bind, label,

detect, and/or cleave nucleic acids. For example, Vlassov, V.V. *et al.* (1986) *Nucleic Acids Res* 14:4065-4076, describe covalent bonding of a single-stranded DNA fragment with alkylating derivatives of nucleotides complementary to target sequences. A report of similar work by the same group is that by Knorre *et al.* (1985) *Biochimie* 67:785-789.

5 Iverson and Dervan also showed sequence-specific cleavage of single-stranded DNA mediated by incorporation of a modified nucleotide which was capable of activating cleavage ((1987) *J Am Chem Soc* 109:1241-1243). Meyer *et al.* (1989) *J Am Chem Soc* 111:8517-8519, effect covalent crosslinking to a target nucleotide using an alkylating agent complementary to the single-stranded target nucleotide sequence. A photoactivated  
10 crosslinking to single-stranded oligonucleotides mediated by psoralen was disclosed by Lee *et al.* (1988) *Biochemistry* 27:3197-3203. Use of crosslinking in triple-helix forming probes was also disclosed by Home *et al.* (1990) *J Am Chem Soc* 112:2435-2437. Use of N4, N4-ethanocytosine as an alkylating agent to crosslink to single-stranded oligonucleotides has also been described by Webb *et al.* (1986) *J Am Chem Soc* (1986)  
15 108:2764-2765; (1986) *Nucleic Acids Res* 14:7661-7674; Feteritz *et al.* (1991) *J. Am. Chem. Soc.* 113:4000. Various compounds to bind, detect, label, and/or cleave nucleic acids are known in the art. See, for example, U.S. Patent Nos. 5,543,507; 5,672,593; 5,484,908; 5,256,648; and 5,681,941.

## 20 Proteins

The isolated proteins of the present invention comprise a polypeptide having at least 10 amino acids encoded by any one of the polynucleotides of the present invention as discussed more fully, above, or polypeptides which are conservatively modified variants thereof. The proteins of the present invention or variants thereof can comprise  
25 any number of contiguous amino acid residues from a polypeptide of the present invention, wherein that number is selected from the group of integers consisting of from 10 to the number of residues in a full-length polypeptide of the present invention. Optionally, this subsequence of contiguous amino acids is at least 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 37, 38, 39, or 40 amino acids in

length, often at least 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 amino acids in length. Further, the number of such subsequences can be any integer selected from the group consisting of from 1 to 20, such as 2, 3, 4, or 5.

As those of skill will appreciate, the present invention includes catalytically active polypeptides of the present invention (i.e., enzymes). Catalytically active polypeptides have a specific activity of at least 20%, 30%, or 40%, and preferably at least 50%, 60%, or 70%, and most preferably at least 80%, 90%, or 95% that of the native (non-synthetic), endogenous polypeptide. Further, the substrate specificity ( $k_{cat}/K_m$ ) is optionally substantially similar to the native (non-synthetic), endogenous polypeptide. Typically, the  $K_m$  will be at least 30%, 40%, or 50%, that of the native (non-synthetic), endogenous polypeptide; and more preferably at least 60%, 70%, 80%, or 90%. Methods of assaying and quantifying measures of enzymatic activity and substrate specificity ( $k_{cat}/K_m$ ), are well known to those of skill in the art.

Generally, the proteins of the present invention will, when presented as an immunogen, elicit production of an antibody specifically reactive to a polypeptide of the present invention. Further, the proteins of the present invention will not bind to antisera raised against a polypeptide of the present invention, which has been fully immunosorbed with the same polypeptide. Immunoassays for determining binding are well known to those of skill in the art. A method of immunoassay is a competitive immunoassay as discussed, *infra*. Thus, the proteins of the present invention can be employed as immunogens for constructing antibodies immunoreactive to a protein of the present invention for such exemplary utilities as immunoassays or protein purification techniques.

By "variant" protein is intended a protein derived from the native protein by deletion (so-called truncation) or addition of one or more amino acids to the N-terminal and/or C-terminal end of the native protein; deletion or addition of one or more amino acids at one or more sites in the native protein; or substitution of one or more amino acids at one or more sites in the native protein. Variant proteins encompassed by the present invention are biologically active, that is they continue to possess the desired biological



activity of the native protein, that is, a pathogenesis-related activity as described herein. Such variants may result from, for example, genetic polymorphism or from human manipulation. Biologically active variants of a native PR1-C10 protein of the invention will have at least about 60%, 61%, 65%, 70%, generally at least about 75%, 80%, 85%,  
5 preferably at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, and more preferably at least about 98%, 99% or more sequence identity to the amino acid sequence for the native protein as determined by sequence alignment programs described elsewhere herein using default parameters. A biologically active variant of a protein of the invention may differ from that protein by as few as 1-15 amino acid residues, as few  
10 as 1-10, such as 6-10, as few as 5, as few as 4, 3, 2, or even 1 amino acid residue.

The proteins of the invention may be altered in various ways including amino acid substitutions, deletions, truncations, and insertions. Methods for such manipulations are generally known in the art. For example, amino acid sequence variants of the PR1-C10 proteins can be prepared by mutations in the DNA. Methods for mutagenesis and  
15 nucleotide sequence alterations are well known in the art. See, for example, Kunkel (1985) *Proc. Natl. Acad. Sci. USA* 82:488-492; Kunkel *et al.* (1987) *Methods in Enzymol.* 154:367-382; U.S. Patent No. 4,873,192; Walker and Gaastra, eds. (1983) *Techniques in Molecular Biology* (MacMillan Publishing Company, New York) and the references cited therein. Guidance as to appropriate amino acid substitutions that do not affect biological  
20 activity of the protein of interest may be found in the model of Dayhoff *et al.* (1978) *Atlas of Protein Sequence and Structure* (Natl. Biomed. Res. Found., Washington, D.C.), herein incorporated by reference. Conservative substitutions, such as exchanging one amino acid with another having similar properties, may be preferable.

## 25 Expression of Proteins in Host Cells

Using the nucleic acids of the present invention, one may express a protein of the present invention in a recombinantly engineered cell such as bacteria, yeast, insect, mammalian, or preferably plant cells. The cells produce the protein in a non-natural

condition (e.g., in quantity, composition, location, and/or time), because they have been genetically altered through human intervention to do so.

It is expected that those of skill in the art are knowledgeable in the numerous expression systems available for expression of a nucleic acid encoding a protein of the present invention. No attempt to describe in detail the various methods known for the expression of proteins in prokaryotes or eukaryotes will be made.

In brief summary, the expression of isolated nucleic acids encoding a protein of the present invention will typically be achieved by operably linking, for example, the DNA or cDNA to a promoter (which is either constitutive or regulatable), followed by incorporation into an expression vector. The vectors can be suitable for replication and integration in either prokaryotes or eukaryotes. Typical expression vectors contain transcription and translation terminators, initiation sequences, and promoters useful for regulation of the expression of the DNA encoding a protein of the present invention. To obtain high level expression of a cloned gene, it is desirable to construct expression vectors which contain, at the minimum, a strong promoter to direct transcription, a ribosome binding site for translational initiation, and a transcription/translation terminator. One of skill would recognize that modifications could be made to a protein of the present invention without diminishing its biological activity. Some modifications may be made to facilitate the cloning, expression, or incorporation of the targeting molecule into a fusion protein. Such modifications are well known to those of skill in the art and include, for example, a methionine added at the amino terminus to provide an initiation site, or additional amino acids (e.g., poly His) placed on either terminus to create conveniently located purification sequences. Restriction sites or termination codons can also be introduced.

#### *A. Expression in Prokaryotes*

Prokaryotic cells may be used as hosts for expression. Prokaryotes most frequently are represented by various strains of *E. coli*; however, other microbial strains may also be used. Commonly used prokaryotic control sequences which are defined

herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding sequences, include such commonly used promoters as the beta lactamase (penicillinase) and lactose (lac) promoter systems (Chang *et al.* (1977) *Nature* 198:1056), the tryptophan (trp) promoter system (Goeddel *et al.* (1980) *Nucleic Acids Res.* 8:4057) and the lambda derived P L promoter and N-gene ribosome binding site (Shimatake *et al.* (1981) *Nature* 292:128). The inclusion of selection markers in DNA vectors transfected in *E coli.* is also useful. Examples of such markers include genes specifying resistance to ampicillin, tetracycline, or chloramphenicol.

The vector is selected to allow introduction into the appropriate host cell.

10 Bacterial vectors are typically of plasmid or phage origin. Appropriate bacterial cells are infected with phage vector particles or transfected with naked phage vector DNA. If a plasmid vector is used, the bacterial cells are transfected with the plasmid vector DNA. Expression systems for expressing a protein of the present invention are available using *Bacillus sp.* and *Salmonella* (Palva *et al.* (1983) *Gene* 22: 229-235; Mosbach *et al.* (1983) 15 *Nature* 302:543-545).

#### *B. Expression in Eukaryotes*

A variety of eukaryotic expression systems such as yeast, insect cell lines, plant and mammalian cells, are known to those of skill in the art. As explained briefly below, a 20 polynucleotide of the present invention can be expressed in these eukaryotic systems. In some embodiments, transformed/transfected plant cells, as discussed *infra*, are employed as expression systems for production of the proteins of the instant invention.

Synthesis of heterologous proteins in yeast is well known. Sherman *et al.* (1982) *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory is a well recognized work 25 describing the various methods available to produce the protein in yeast. Two widely utilized yeasts for production of eukaryotic proteins are *Saccharomyces cerevisiae* and *Pichia pastoris*. Vectors, strains, and protocols for expression in *Saccharomyces* and *Pichia* are known in the art and available from commercial suppliers (e.g., Invitrogen). Suitable vectors usually have expression control sequences, such as promoters, including

3-phosphoglycerate kinase or alcohol oxidase, and an origin of replication, termination sequences and the like as desired.

A protein of the present invention, once expressed, can be isolated from yeast by lysing the cells and applying standard protein isolation techniques to the lysates. The  
5 monitoring of the purification process can be accomplished by using Western blot techniques, radioimmunoassay, or other standard immunoassay techniques.

The sequences encoding proteins of the present invention can also be ligated to various expression vectors for use in transfecting cell cultures of, for instance, mammalian, insect, or plant origin. Illustrative cell cultures useful for the production of  
10 the peptides are mammalian cells. Mammalian cell systems often will be in the form of minelayers of cells although mammalian cell suspensions may also be used. A number of suitable host cell lines capable of expressing intact proteins have been developed in the art, and include the HEK293, BHK21, and CHO cell lines. Expression vectors for these cells can include expression control sequences, such as an origin of replication, a  
15 promoter (*e.g.* the CMV promoter, a HSV *tk* promoter or *pgk* (phosphoglycerate kinase) promoter), an enhancer (Queen *et al.* (1986) *Immunol. Rev.* 89:49), and necessary processing information sites, such as ribosome binding sites, RNA splice sites, polyadenylation sites (*e.g.*, an SV40 large T Ag poly A addition site), and transcriptional terminator sequences. Other animal cells useful for production of proteins of the present  
20 invention are available, for instance, from the American Type Culture Collection.

Appropriate vectors for expressing proteins of the present invention in insect cells are usually derived from the SF9 baculovirus. Suitable insect cell lines include mosquito larvae, silkworm, armyworm, moth and *Drosophila* cell lines such as a Schneider cell line (See, Schneider, (1987) *J. Embryol. Exp. Morphol.* 27:353-365).

25 As with yeast, when higher animal or plant host cells are employed, polyadenylation or transcription terminator sequences are typically incorporated into the vector. An example of a terminator sequence is the polyadenylation sequence from the bovine growth hormone gene. Sequences for accurate splicing of the transcript may also be included. An example of a splicing sequence is the VP1 intron from SV40 (Sprague *et*

*al.* (1983) *J. Virol.* 45:773-781). Additionally, gene sequences to control replication in the host cell may be incorporated into the vector such as those found in bovine papilloma virus type-vectors. Saveria-Campo (1985) *DNA Cloning Vol. II a Practical Approach*, D.M. Glover, IRL Press, Arlington, Virginia pp. 213-238.

5

#### Transfection/Transformation of Cells

The method of transformation/transfection is not critical to the instant invention; various methods of transformation or transfection are currently available. As newer methods are available to transform crops or other host cells they may be directly applied.

10 Accordingly, a wide variety of methods have been developed to insert a DNA sequence into the genome of a host cell to obtain the transcription and/or translation of the sequence to effect phenotypic changes in the organism. Thus, any method, which provides for effective transformation/transfection may be employed.

#### 15 *A. Plant Transformation*

The genes of the present invention can be used to transform any plant. In this manner, genetically modified plants, plant cells, plant tissue, seed, and the like can be obtained. Transformation protocols as well as protocols for introducing nucleotide sequences into plants may vary depending on the type of plant or plant cell, i.e., monocot or dicot, targeted for transformation. Suitable methods of introducing nucleotide sequences into plant cells and subsequent insertion into the plant genome include microinjection (Crossway *et al.* (1986) *Biotechniques* 4:320-334), electroporation (Riggs *et al.* (1986) *Proc. Natl. Acad. Sci. USA* 83:5602-5606, *Agrobacterium*-mediated transformation (Townsend *et al.* U.S. Patent No. 5,563,055; Zhao *et al.* U.S. Patent No. 20 5,981,840), direct gene transfer (Paszkowski *et al.* (1984) *EMBO J.* 3:2717-2722), and ballistic particle acceleration (see, for example, Sanford *et al.*, U.S. Patent No. 4,945,050; Tomes *et al.*, U.S. Patent No. 5,879,918; Tomes *et al.*, U.S. Patent No. 5,886,244; Bidney *et al.*, U.S. Patent No. 5,932,782; Tomes *et al.* (1995) "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment," in *Plant Cell, Tissue, and Organ Culture*:

25

*Fundamental Methods*, ed. Gamborg and Phillips (Springer-Verlag, Berlin); and McCabe *et al.* (1988) *Biotechnology* 6:923-926). Also see Weissinger *et al.* (1988) *Ann. Rev. Genet.* 22:421-477; Sanford *et al.* (1987) *Particulate Science and Technology* 5:27-37 (onion); Christou *et al.* (1988) *Plant Physiol.* 87:671-674 (soybean); McCabe *et al.* (1988) *Bio/Technology* 6:923-926 (soybean); Finer and McMullen (1991) *In Vitro Cell Dev. Biol.* 27P:175-182 (soybean); Singh *et al.* (1998) *Theor. Appl. Genet.* 96:319-324 (soybean); Datta *et al.* (1990) *Biotechnology* 8:736-740 (rice); Klein *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:4305-4309 (maize); Klein *et al.* (1988) *Biotechnology* 6:559-563 (maize); Tomes, U.S. Patent No. 5,240,855; Busing *et al.*, U.S. Patent Nos. 5,322,783 and 5,324,646; Tomes *et al.* (1995) "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment," in *Plant Cell, Tissue, and Organ Culture: Fundamental Methods*, ed. Gamborg (Springer-Verlag, Berlin) (maize); Klein *et al.* (1988) *Plant Physiol.* 91:440-444 (maize); Fromm *et al.* (1990) *Biotechnology* 8:833-839 (maize); Hooykaas-Van Slogteren *et al.* (1984) *Nature (London)* 311:763-764; Bowen *et al.*, U.S. Patent No. 5,736,369 (cereals); Bytebier *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:5345-5349 (Liliaceae); De Wet *et al.* (1985) in *The Experimental Manipulation of Ovule Tissues*, ed. Chapman *et al.* (Longman, New York), pp. 197-209 (pollen); Kaeppeler *et al.* (1990) *Plant Cell Reports* 9:415-418 and Kaeppeler *et al.* (1992) *Theor. Appl. Genet.* 84:560-566 (whisker-mediated transformation); D'Halluin *et al.* (1992) *Plant Cell* 4:1495-1505 (electroporation); Li *et al.* (1993) *Plant Cell Reports* 12:250-255 and Christou and Ford (1995) *Annals of Botany* 75:407-413 (rice); Osjoda *et al.* (1996) *Nature Biotechnology* 14:745-750 (maize via *Agrobacterium tumefaciens*); all of which are herein incorporated by reference.

The methods of the invention involve introducing a nucleotide construct into a plant. By "introducing" is intended presenting to the plant the nucleotide construct in such a manner that the construct gains access to the interior of a cell of the plant. The methods of the invention do not depend on a particular method for introducing a nucleotide construct to a plant, only that the nucleotide construct gains access to the interior of at least one cell of the plant. Methods for introducing nucleotide constructs

into plants are known in the art including, but not limited to, stable transformation methods, transient transformation methods, and virus-mediated methods.

By “stable transformation” is intended that the nucleotide construct introduced into a plant integrates into the genome of the plant and is capable of being inherited by progeny thereof. By “transient transformation” is intended that a nucleotide construct introduced into a plant does not integrate into the genome of the plant.

The nucleotide constructs of the invention may be introduced into plants by contacting plants with a virus or viral nucleic acids. Generally, such methods involve incorporating a nucleotide construct of the invention within a viral DNA or RNA molecule. It is recognized that a PR1-C10 of the invention may be initially synthesized as part of a viral polyprotein, which later may be processed by proteolysis *in vivo* or *in vitro* to produce the desired recombinant protein. Further, it is recognized that promoters of the invention also encompass promoters utilized for transcription by viral RNA polymerases. Methods for introducing nucleotide constructs into plants and expressing a protein encoded therein, involving viral DNA or RNA molecules, are known in the art. See, for example, U.S. Patent Nos. 5,889,191, 5,889,190, 5,866,785, 5,589,367 and 5,316,931; herein incorporated by reference.

The cells, which have been transformed, may be grown into plants in accordance with conventional ways. See, for example, McCormick *et al.* (1986) *Plant Cell Reports*, 5:81-84. These plants may then be grown, and either pollinated with the same transformed strain or different strains, and the resulting hybrid having the desired phenotypic characteristic identified. Two or more generations may be grown to ensure that the subject phenotypic characteristics are stably maintained and inherited and then seeds harvested to ensure expression of the desired phenotype or other property has been achieved. One of skill will recognize that after the recombinant expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of number of standard breeding techniques can be used, depending upon the species to be crossed.

In vegetatively propagated crops, mature transgenic plants can be propagated by the taking of cuttings or by tissue culture techniques to produce multiple identical plants. Selection of desirable transgenics is made and new varieties are obtained and propagated vegetatively for commercial use. In seed propagated crops, mature transgenic plants can be self crossed to produce a homozygous inbred plant. The inbred plant produces seed containing the newly introduced heterologous nucleic acid. These seeds can be grown to produce plans that would produce the selected phenotype.

Parts obtained from the regenerated plant, such as flowers, seeds, leaves, branches, fruit, and the like are included in the invention, provided that these parts comprise cells comprising the isolated nucleic acid of the present invention. Progeny and variants, and mutants of the regenerated plants are also included within the scope of the invention, provided that these parts comprise the introduced nucleic acid sequences.

Another embodiment is a transgenic plant that is homozygous for the added heterologous nucleic acid; i.e., a transgenic plant that contains two added nucleic acid sequences, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be obtained by sexually mating(selfing) a heterozygous transgenic plant that contains a single added heterologous nucleic acid, germinating some of the seed produced and analyzing the resulting plants produced for altered expression of a polynucleotide of the present invention relative to a control plant (i.e., native, non-transgenic). Backcrossing to a parental plant and out-crossing with a non-transgenic plant are also contemplated.

#### *B. Transfection of Prokaryotes, Lower Eukaryotes, and Animal Cells*

Animal and lower eukaryotic (e.g., yeast) host cells are competent or rendered competent for transfection by various means. There are several well-known methods of introducing DNA into animal cells. These include: calcium phosphate precipitation, fusion of the recipient cells with bacterial protoplasts containing the DNA, treatment of the recipient cells with liposomes containing the DNA, DEAE dextrin, electroporation, biolistics, and micro-injection of the DNA directly into the cells. The transfected cells



are cultured by means well known in the art. Kuchler (1997) *Biochemical Methods in Cell Culture and Virology*, Dowden, Hutchinson and Ross, Inc.

### Synthesis of Proteins

- 5           The proteins of the present invention can be constructed using non-cellular synthetic methods. Solid phase synthesis of proteins of less than about 50 amino acids in length may be accomplished by attaching the C-terminal amino acid of the sequence to an insoluble support followed by sequential addition of the remaining amino acids in the sequence. Techniques for solid phase synthesis are described by Barany and Mayfield
- 10   Solid-Phase Peptide Synthesis, pp. 3-284 in *the Peptides: Analysis, Synthesis, Biology. Vol. 2: Special Methods in Peptide Synthesis, Part A*; Merrifield, et al., (1963) *J. Am. Chem. Soc.* 85: 2149-2156 and Stewart et al. (1984) *Solid Phase Peptide Synthesis*, 2<sup>nd</sup> ed., Pierce Chem. Co., Rockford, Ill. Proteins of greater length may be synthesized by condensation of the amino and carboxy termini of shorter fragments. Methods of
- 15   forming peptide bonds by activation of a carboxy terminal end (e.g., by the use of the coupling reagent N,N'-dicyclohexylcarbodiimide)) is known to those of skill in the art.

### Purification of Proteins

- 20           The proteins of the present invention may be purified by standard techniques well known to those of skill in the art. Recombinantly produced proteins of the present invention can be directly expressed or expressed as a fusion protein. The recombinant protein is purified by a combination of cell lysis (e.g., sonication, French Press) and affinity chromatography. For fusion products, subsequent digestion of the fusion protein with an appropriate proteolytic enzyme releases the desired recombinant protein.
- 25           The proteins of this invention, recombinant or synthetic, may be purified to substantial purity by standard techniques well known in the art, including detergent solubilization, selective precipitation with such substances as ammonium sulfate, column chromatography, immunopurification methods, and others. See, for instance, Scopes (1982) *Protein Purification: Principles and Practice*, Springer-Verlag: New York;

Deutscher (1990) *Guide to Protein Purification*, Academic Press. For example, antibodies may be raised to the proteins as described herein. Purification from *E. coli* can be achieved following procedures described in U.S. patent No. 4,511,503. The protein may then be isolated from cells expressing the protein and further purified by standard protein chemistry techniques as described herein. Detection of the expressed protein is achieved by methods known in the art and include, for example, radioimmunoassays, Western blotting techniques, or immunoprecipitation.

#### Modulating Polypeptide Levels and/or Composition

The present invention further provides a method for modulating (i.e., increasing or decreasing) the concentration or composition of the polypeptides of the present invention in a plant or part thereof. Increasing or decreasing the concentration and/or the composition (i.e., the ratio of the polypeptides of the present invention) in a plant can effect modulation. The method comprises introducing into a plant cell a recombinant expression cassette comprising a polynucleotide of the present invention as described above to obtain a transformed plant cell, culturing the transformed plant cell under plant cell growing conditions, and inducing or repressing expression of a polynucleotide of the present invention in the plant for a time sufficient to modulate concentration and/or composition in the plant or plant part.

In some embodiments, the content and/or composition of polypeptides of the present invention in a plant may be modulated by altering, *in vivo* or *in vitro*, the promoter of a gene to up- or down- regulate gene expression. In some embodiments, the coding regions of native genes of the present invention can be altered via substitution, addition, insertion, or deletion to decrease activity of the encoded enzyme. See, e.g., Kmiec, U.S. Patent 5,565,350; Zarling *et al.*, PCT/US93/03868. And in some embodiments, an isolated nucleic acid (e.g., a vector) comprising a promoter sequence is transfected into a plant cell. Subsequently, a plant cell comprising the promoter operably linked to a polynucleotide of the present invention is selected for by means known to those of skill in the art such as, but not limited to, Southern blot, DNA

sequencing, or PCR analysis using primers specific to the promoter and to the gene and detecting amplicons produced therefrom. A plant or plant part altered or modified by the foregoing embodiments is grown under plant forming conditions for a time sufficient to modulate the concentration and/or composition of polypeptides of the present invention  
5 in the plant. Plant forming conditions are well known in the art and discussed briefly, *supra*.

In general, the concentration of the compositions of the invention is increased or decreased by at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% relative to a native control plant, plant part, or cell lacking the aforementioned recombinant  
10 expression cassette. Modulation in the present invention may occur during and/or subsequent to growth of the plant to the desired stage of development. Modulating nucleic acid expression temporally and/or in particular tissues can be controlled by employing the appropriate promoter operably linked to a polynucleotide of the present invention in, for example, sense or antisense orientation as discussed in greater detail,  
15 *supra*. Induction of expression of a polynucleotide of the present invention can also be controlled by exogenous administration of an effective amount of inducing compound. Inducible promoters and inducing compounds, which activate expression from these promoters, are well known in the art. In particular embodiments, the polypeptides of the present invention are modulated in monocots, particularly maize.

#### Molecular Markers

The present invention provides a method of genotyping a plant comprising a polynucleotide of the present invention. Optionally, the plant is a monocot, such as maize or sorghum. Genotyping provides a means of distinguishing homologs of a  
25 chromosome pair and can be used to differentiate segregants in a plant population. Molecular marker methods can be used for phylogenetic studies, characterizing genetic relationships among crop varieties, identifying crosses or somatic hybrids, localizing chromosomal segments affecting monogenic traits, map based cloning, and the study of quantitative inheritance. See, e.g., Clark *et al.* (1997) *Plant Molecular Biology: A*

*Laboratory Manual*, Chapter 7, Springer-Verlag, Berlin. For molecular marker methods, see generally, Paterson *et al.* (1996) (Chapter 2) *Genome Mapping in plants* Andrew H. Paterson, Academic Press/R.G. Lands Company, Austin, Texas, pp. 7-21.

The particular method of genotyping in the present invention may employ any number of molecular marker analytic techniques such as, but not limited to, restriction fragment length polymorphism's (RFLPs). RFLPs are the product of allelic differences between DNA restriction fragments resulting from nucleotide sequence variability. As is well known to those of skill in the art, RFLPs are typically detected by extraction of genomic DNA and digestion with a restriction enzyme. Generally, the resulting fragments are separated according to size and hybridized with a probe; single copy probes are preferred. Restriction fragments from homologous chromosomes are revealed. Differences in fragment size among alleles represent an RFLP. Thus, the present invention further provides a means to follow segregation of a gene or nucleic acid of the present invention as well as chromosomal sequences genetically linked to these genes or nucleic acids using such techniques as RFLP analysis. Linked chromosomal sequences are within 50 centiMorgans (cM), often within 40 or 30 cM, preferably within 20 or 10 cM, more preferably within 5, 3, 2, or 1 cM of a gene of the present invention.

In the present invention, the nucleic acid probes employed for molecular marker mapping of plant nuclear genomes selectively hybridize, under selective hybridization conditions, to a gene encoding a polynucleotide of the present invention. In one embodiment, the probes are selected from polynucleotides of the present invention. Typically, these probes are cDNA probes or restriction enzyme treated (e.g., *Pst I*) genomic clones. The length of the probes is discussed in greater detail, *supra*, but is typically at least 15 bases in length, more preferably at least 20, 25, 30, 35, 40, or 50 bases in length. Generally, however, the probes are less than about 1 kilobase in length. Preferably, the probes are single copy probes that hybridize to a unique locus in haploid chromosome compliment. Some exemplary restriction enzymes employed in RFLP mapping are *EcoRI*, *EcoRV*, and *SstI*. As used herein the term "restriction enzyme"

includes reference to a composition that recognizes and, alone or in conjunction with another composition, cleaves at a specific nucleotide sequence.

The method of detecting an RFLP comprises the steps of (a) digesting genomic DNA of a plant with a restriction enzyme; (b) hybridizing a nucleic acid probe, under  
5 selective hybridization conditions, to a sequence of a polynucleotide of the present of said genomic DNA; (c) detecting therefrom a RFLP. Other methods of differentiating polymorphic (allelic) variants of polynucleotides of the present invention can be had by utilizing molecular marker techniques well known to those of skill in the art including such techniques as: 1) single stranded conformation analysis (SSCA); 2) denaturing  
10 gradient gel electrophoresis (DGGE); 3) RNase protection assays; 4) allele-specific oligonucleotides (ASOs); 5) the use of proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein; and 6) allele-specific PCR. Other approaches based on the detection of mismatches between the two complementary DNA strands include clamped denaturing gel electrophoresis (CDGE); heteroduplex analysis (HA); and  
15 chemical mismatch cleavage (CMC). Thus, the present invention further provides a method of genotyping comprising the steps of contacting, under stringent hybridization conditions, a sample suspected of comprising a polynucleotide of the present invention with a nucleic acid probe. Generally, the sample is a plant sample, preferably, a sample suspected of comprising a maize polynucleotide of the present invention (e.g., gene,  
20 mRNA). The nucleic acid probe selectively hybridizes, under stringent conditions, to a subsequence of a polynucleotide of the present invention comprising a polymorphic marker. Selective hybridization of the nucleic acid probe to the polymorphic marker nucleic acid sequence yields a hybridization complex. Detection of the hybridization complex indicates the presence of that polymorphic marker in the sample. In another  
25 embodiment, the nucleic acid probe comprises a polynucleotide of the present invention.

#### UTRs and Codon Preference

In general, translational efficiency has been found to be regulated by specific sequence elements in the 5' non-coding or untranslated region (5' UTR) of the RNA.

Positive sequence motifs include translational initiation consensus sequences (Kozak (1987) *Nucleic Acids Res* 15:8125) and the 7-methylguanosine cap structure (Drummond *et al.* (1985) *Nucleic Acids Res.* 13:7375). Negative elements include stable intramolecular 5' UTR stem-loop structures (Muesing *et al.* (1987) *Cell* 48:691) and

5 AUG sequences or short open reading frames preceded by an appropriate AUG in the 5' UTR (Kozak, *supra*, Rao *et al.* (1988) *Mol. and Cell. Biol.* 8:284). Accordingly, the present invention provides 5' and/or 3' UTR regions for modulation of translation of heterologous coding sequences.

Further, the polypeptide-encoding segments of the polynucleotides of the present

10 invention can be modified to alter codon usage. Altered codon usage can be employed to alter translational efficiency and/or to optimize the coding sequence for expression in a desired host such as to optimize the codon usage in a heterologous sequence for expression in maize. Codon usage in the coding regions of the polynucleotides of the present invention can be analyzed statistically using commercially available software

15 packages such as "Codon Preference" available from the University of Wisconsin Genetics Computer Group (see Devereaux *et al.* (1984) *Nucleic Acids Res.* 12:387-395) or MacVector 4.1 (Eastman Kodak Co., New Haven, Conn.). Thus, the present invention provides a codon usage frequency characteristic of the coding region of at least one of the polynucleotides of the present invention. The number of polynucleotides that can be used

20 to determine a codon usage frequency can be any integer from 1 to the number of polynucleotides of the present invention as provided herein. Optionally, the polynucleotides will be full-length sequences. An exemplary number of sequences for statistical analysis can be at least 1, 5, 10, 20, 50, or 100.

## 25 Sequence Shuffling

The present invention provides methods for sequence shuffling using polynucleotides of the present invention, and compositions resulting therefrom. Sequence shuffling is described in PCT publication No. WO 96/19256. See also, Zhang *et al.* (1997) *Proc. Natl. Acad. Sci. USA* 94:4504-4509 (1997). Generally, sequence

shuffling provides a means for generating libraries of polynucleotides having a desired characteristic, which can be selected or screened for. Libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides which comprise sequence regions, which have substantial identity and can be

5 homologously recombined *in vitro* or *in vivo*. The population of sequence-recombined polynucleotides comprises a subpopulation of polynucleotides which possess desired or advantageous characteristics and which can be selected by a suitable selection or screening method. The characteristics can be any property or attribute capable of being selected for or detected in a screening system, and may include properties of: an encoded  
10 protein, a transcriptional element, a sequence controlling transcription, RNA processing, RNA stability, chromatin conformation, translation, or other expression property of a gene or transgene, a replicative element, a protein-binding element, or the like, such as any feature which confers a selectable or detectable property. In some embodiments, the selected characteristic will be a decreased  $K_m$  and/or increased  $K_{cat}$  over the wild-type  
15 protein as provided herein. In other embodiments, a protein or polynucleotide generated from sequence shuffling will have a ligand binding affinity greater than the non-shuffled wild-type polynucleotide. The increase in such properties can be at least 110%, 120%, 130%, 140%, or at least 150% of the wild-type value.

## 20 Generic and Consensus Sequences

Polynucleotides and polypeptides of the present invention further include those having: (a) a generic sequence of at least two homologous polynucleotides or polypeptides, respectively, of the present invention; and, (b) a consensus sequence of at least three homologous polynucleotides or polypeptides, respectively, of the present  
25 invention. The generic sequence of the present invention comprises each species of polypeptide or polynucleotide embraced by the generic polypeptide or polynucleotide, sequence, respectively. The individual species encompassed by a polynucleotide having an amino acid or nucleic acid consensus sequence can be used to generate antibodies or produce nucleic acid probes or primers to screen for homologs in other species, genera,

families, orders, classes, phylums, or kingdoms. For example, a polynucleotide having a consensus sequences from a gene family of *Zea mays* can be used to generate antibody or nucleic acid probes or primers to other *Gramineae* species such as wheat, rice, or sorghum. Alternatively, a polynucleotide having a consensus sequence generated from orthologous genes can be used to identify or isolate orthologs of other taxa. Typically, a polynucleotide having a consensus sequence will be at least 9, 10, 15, 20, 25, 30, or 40 amino acids in length, or 20, 30, 40, 50, 100, or 150 nucleotides in length. As those of skill in the art are aware, a conservative amino acid substitution can be used for amino acids, which differ amongst aligned sequence but are from the same conservative amino substitution group as discussed above. Optionally, no more than 1 or 2 conservative amino acids are substituted for each 10 amino acid length of consensus sequence.

Similar sequences used for generation of a consensus or generic sequence include any number and combination of allelic variants of the same gene, orthologous, or paralogous sequences as provided herein. Optionally, similar sequences used in generating a consensus or generic sequence are identified using the BLAST algorithm's smallest sum probability (P(N)). Various suppliers of sequence-analysis software are listed in chapter 7 of Ausubel *et al. Current Protocols in Molecular Biology*, Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc. (Supplement 30). A polynucleotide sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less then about 0.1, more preferably less than about 0.01, or 0.001, and most preferably less than about 0.0001, or 0.00001. Similar polynucleotides can be aligned and a consensus or generic sequence generated using multiple sequence alignment software available from a number of commercial suppliers such as the Genetics Computer Group's (Madison, WI) PILEUP software, Vector NTI's (North Bethesda, MD) ALIGNX, or Genecode's (Ann Arbor, MI) SEQUENCER. Conveniently, default parameters of such software can be used to generate consensus or generic sequences.



Although the present invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practices within the scope of the appended claims.

5

## EXPERIMENTAL

### Example 1: Isolation of PR1-C10

A partial sequence of the maize PR1-C10 polynucleotide was discovered by sequencing a clone derived from a randomly selected maize cDNA library in yeast. The cDNA library was made from mRNA isolated from maize cells. The maize cells were treated with water or  $1 \times 10^6$  spores/ml of *Fusarium moniliforme*. Cells were harvested two and six hours after treatment. Total RNA was isolated using Tri-Reagent<sup>TM</sup> and mRNA was isolated using PolyAtract<sup>TM</sup> (Promega). Zap-cDNA synthesis kit (Stratagene) was used to prepare cDNA, which was cloned into HybriZap<sup>®</sup> (Stratagene). The primary library was amplified and phagemid was excised from the secondary library. The phagemid prep was amplified in XL0LR cells and purified (Qiagen) to prepare library DNA for transformation into yeast. All library manipulations were performed according to the HybriZap<sup>®</sup> manual. The partial PR1-C10 clone was then used to probe a second lambda library, prepared in the same manner as described for the first lambda library. Ten identical, full length clones were isolated and sequenced. The sequence of the PR1-C10 polynucleotide and related polypeptide are shown in SEQ ID NOS:1 and 2, respectively. The coding sequence of SEQ ID NO:1 (nucleotides 63-674 of SEQ ID NO:1) that encodes the polypeptide set forth in SEQ ID NO:2 is set forth in SEQ ID NO:3.

### Example 2: Identification and Characterization of PR1-C10

Gene identities were determined by conducting BLAST (Basic Local Alignment Search Tool; Altschul, *et al.* (1993) *J. Mol. Biol.* 215:403-410; see also

www.ncbi.nlm.nih.gov/BLAST/) searches under default parameters for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm. The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish *et al.* (1993) *Nature Genetics* 3:266-272) provided by the NCBI. In some cases, the sequencing data from two or more clones containing overlapping segments of DNA were used to construct contiguous DNA sequences. The maize PR1-C10 polynucleotide shares an overall homology to other maize PR1 genes such as PR1-#83 (60.5% identity), PR1-#52 (60.3% identity), PR1-#93 (58.4% identity), PR1-#81 (54.6% identity) and PR1-#70 (50.0% identity). Please see, PCT published application number WO 99/43819, published September 9, 1999, for a description of the maize PR1 genes. In addition, the maize PR1-C10 polynucleotide shows homology to two unknown rice partial cDNA sequences having Genebank Accession numbers AU029886 and AU029887.

### Example 3: Expression of PR1-C10

To determine the expression pattern of the native PR1-C10 gene, Northern analysis was performed according to Sambrook *et al. supra*. Maize GS3 suspension cell cultures were treated with water or  $1 \times 10^6$  spores/ml of *Fusarium moniliforme* and harvested at 2 hours and 6 hours after treatment. The Northern Blot analysis showed that PR1-C10 is a fairly abundant transcript in maize GS3 cells, but it is not induced by either elicitor or spore treatment. In healthy, green house-grown HG11 plants, PR1-C10 is expressed at low levels in 8 days after pollination kernels, V12 leaf-blade joints (i.e. auricles and ligules), green tassels and shedding tassels. The PR1-C10 mRNA level is abundant in R1 (silking stage) leaf-blade joints. At the level of detection provided by

Northern blot hybridization, PR1-C10 is not expressed in 10 day-old or mature leaf blades, pith, roots, husks, silks, leaf sheaths or immature ears.

#### Example 4: Transformation and Regeneration of Transgenic Plants

5           Immature maize embryos from greenhouse donor plants are bombarded with a plasmid containing the PR1-C10 operably linked to a ubiquitin promoter and the selectable marker gene PAT (Wohlleben *et al.* (1988) *Gene* 70:25-37), which confers resistance to the herbicide Bialaphos. Alternatively, the selectable marker gene is provided on a separate plasmid. Transformation is performed as follows. Media recipes  
10 follow below.

#### Preparation of Target Tissue

15           The ears are husked and surface sterilized in 30% Clorox bleach plus 0.5% Micro detergent for 20 minutes, and rinsed two times with sterile water. The immature embryos are excised and placed embryo axis side down (scutellum side up), 25 embryos per plate, on 560Y medium for 4 hours and then aligned within the 2.5-cm target zone in preparation for bombardment.

#### Preparation of DNA

20           A plasmid vector comprising the PR1-C10 operably linked to a ubiquitin promoter is made. This plasmid DNA plus plasmid DNA containing a PAT selectable marker is precipitated onto 1.1  $\mu\text{m}$  (average diameter) tungsten pellets using a  $\text{CaCl}_2$  precipitation procedure as follows:

- 25           100  $\mu\text{l}$  prepared tungsten particles in water  
            10  $\mu\text{l}$  (1  $\mu\text{g}$ ) DNA in Tris EDTA buffer (1  $\mu\text{g}$  total DNA)  
            100  $\mu\text{l}$  2.5 M  $\text{CaCl}_2$   
            10  $\mu\text{l}$  0.1 M spermidine

Each reagent is added sequentially to the tungsten particle suspension, while maintained on the multitube vortexer. The final mixture is sonicated briefly and allowed to incubate under constant vortexing for 10 minutes. After the precipitation period, the tubes are centrifuged briefly, liquid removed, washed with 500 ml 100% ethanol, and centrifuged for 30 seconds. Again the liquid is removed, and 105 µl 100% ethanol is added to the final tungsten particle pellet. For particle gun bombardment, the tungsten/DNA particles are briefly sonicated and 10 µl spotted onto the center of each macrocarrier and allowed to dry about 2 minutes before bombardment.

#### 10 Particle Gun Treatment

The sample plates are bombarded at level #4 in particle gun #HE34-1 or #HE34-2. All samples receive a single shot at 650 PSI, with a total of ten aliquots taken from each tube of prepared particles/DNA.

#### 15 Subsequent Treatment

Following bombardment, the embryos are kept on 560Y medium for 2 days, then transferred to 560R selection medium containing 3 mg/liter Bialaphos, and subcultured every 2 weeks. After approximately 10 weeks of selection, selection-resistant callus clones are transferred to 288J medium to initiate plant regeneration. Following somatic embryo maturation (2-4 weeks), well-developed somatic embryos are transferred to medium for germination and transferred to the lighted culture room. Approximately 7-10 days later, developing plantlets are transferred to 272V hormone-free medium in tubes for 7-10 days until plantlets are well established. Plants are then transferred to inserts in flats (equivalent to 2.5" pot) containing potting soil and grown for 1 week in a growth chamber, subsequently grown an additional 1-2 weeks in the greenhouse, then transferred to classic 600 pots (1.6 gallon) and grown to maturity. Plants are monitored and scored for PRI-C10 expression.

#### 30 Bombardment and Culture Media

RTA01/2093378v1

Bombardment medium (560Y) comprises 4.0 g/l N6 basal salts (SIGMA C-1416), 1.0 ml/l Eriksson's Vitamin Mix (1000X SIGMA-1511), 0.5 mg/l thiamine HCl, 120.0 g/l sucrose, 1.0 mg/l 2,4-D, and 2.88 g/l L-proline (brought to volume with D-I H<sub>2</sub>O following adjustment to pH 5.8 with KOH); 2.0 g/l Gelrite (added after bringing to volume with D-I H<sub>2</sub>O); and 8.5 mg/l silver nitrate (added after sterilizing the medium and cooling to room temperature). Selection medium (560R) comprises 4.0 g/l N6 basal salts (SIGMA C-1416), 1.0 ml/l Eriksson's Vitamin Mix (1000X SIGMA-1511), 0.5 mg/l thiamine HCl, 30.0 g/l sucrose, and 2.0 mg/l 2,4-D (brought to volume with D-I H<sub>2</sub>O following adjustment to pH 5.8 with KOH); 3.0 g/l Gelrite (added after bringing to volume with D-I H<sub>2</sub>O); and 0.85 mg/l silver nitrate and 3.0 mg/l bialaphos (both added after sterilizing the medium and cooling to room temperature).

Plant regeneration medium (288J) comprises 4.3 g/l MS salts (GIBCO 11117-074), 5.0 ml/l MS vitamins stock solution (0.100 g nicotinic acid, 0.02 g/l thiamine HCL, 0.10 g/l pyridoxine HCL, and 0.40 g/l glycine brought to volume with polished D-I H<sub>2</sub>O) (Murashige and Skoog (1962) *Physiol. Plant.* 15:473), 100 mg/l myo-inositol, 0.5 mg/l zeatin, 60 g/l sucrose, and 1.0 ml/l of 0.1 mM abscisic acid (brought to volume with polished D-I H<sub>2</sub>O after adjusting to pH 5.6); 3.0 g/l Gelrite (added after bringing to volume with D-I H<sub>2</sub>O); and 1.0 mg/l indoleacetic acid and 3.0 mg/l bialaphos (added after sterilizing the medium and cooling to 60°C). Hormone-free medium (272V) comprises 4.3 g/l MS salts (GIBCO 11117-074), 5.0 ml/l MS vitamins stock solution (0.100 g/l nicotinic acid, 0.02 g/l thiamine HCL, 0.10 g/l pyridoxine HCL, and 0.40 g/l glycine brought to volume with polished D-I H<sub>2</sub>O), 0.1 g/l myo-inositol, and 40.0 g/l sucrose (brought to volume with polished D-I H<sub>2</sub>O after adjusting pH to 5.6); and 6 g/l bacto-agar (added after bringing to volume with polished D-I H<sub>2</sub>O), sterilized and cooled to 60° C.

For *Agrobacterium*-mediated transformation of maize with a PR1-C10 sequence of the invention, preferably the method of Zhao is employed (U.S. Patent No. 5,981,840, and PCT patent publication WO98/32326; the contents of which are hereby incorporated by reference). Briefly, immature embryos are isolated from maize and the embryos contacted with a suspension of *Agrobacterium*, where the bacteria are capable of transferring the

PR1-C10 nucleotide sequences of interest to at least one cell of at least one of the immature embryos (step 1: the infection step). In this step the immature embryos are preferably immersed in an *Agrobacterium* suspension for the initiation of inoculation. The embryos are co-cultured for a time with the *Agrobacterium* (step 2: the co-cultivation step). Preferably the immature embryos are cultured on solid medium following the infection step. Following this co-cultivation period an optional "resting" step is contemplated. In this resting step, the embryos are incubated in the presence of at least one antibiotic known to inhibit the growth of *Agrobacterium* without the addition of a selective agent for plant transformants (step 3: resting step). Preferably the immature embryos are cultured on solid medium with antibiotic, but without a selecting agent, for elimination of *Agrobacterium* and for a resting phase for the infected cells. Next, inoculated embryos are cultured on medium containing a selective agent and growing transformed callus is recovered (step 4: the selection step). Preferably, the immature embryos are cultured on solid medium with a selective agent resulting in the selective growth of transformed cells. The callus is then regenerated into plants (step 5: the regeneration step), and preferably calli grown on selective medium are cultured on solid medium to regenerate the plants.

#### Example 5: Soybean Embryo Transformation Example

Soybean embryos are bombarded with a plasmid containing the PR1-C10 nucleotide sequence operably linked to a SCPI promoter as follows. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface-sterilized, immature seeds of the soybean cultivar A2872, are cultured in the light or dark at 26°C on an appropriate agar medium for six to ten weeks. Somatic embryos producing secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos that multiplied as early, globular-staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can be maintained in 35 ml liquid media on a rotary shaker, 150 rpm, at 26°C with fluorescent lights on a 16:8 hour day/night

schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 ml of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein *et al.* (1987) *Nature* (London) 327:70-73, U.S.

- 5 Patent No. 4,945,050). A Du Pont Biolistic PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene that can be used to facilitate soybean transformation is a transgene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell *et al.* (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid  
10 pJR225 (from *E. coli*; Gritz *et al.* (1983) *Gene* 25:179-188), and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The expression cassette comprising the PR1-C10 operably linked to the SCPI promoter can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

- 15 To 50 µl of a 60 mg/ml 1 µm gold particle suspension is added (in order): 5 µl DNA (1 µg/µl), 20 µl spermidine (0.1 M), and 50 µl CaCl<sub>2</sub> (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µl 70% ethanol and resuspended in 40 µl of anhydrous ethanol. The DNA/particle suspension  
20 can be sonicated three times for one second each. Five microliters of the DNA-coated gold particles are then loaded on each macro carrier disk.

- Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are  
25 normally bombarded. Membrane rupture pressure is set at 1100 psi, and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post-bombardment with fresh media containing 50 mg/ml hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post-bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

#### Example 6: Sunflower Meristem Tissue Transformation Prophetic Example

Sunflower meristem tissues are transformed with an expression cassette containing the PR1-C10 operably linked to a SCPI promoter as follows (see also European Patent Number EP 0 486233, herein incorporated by reference, and Malone-Schoneberg *et al.* (1994) *Plant Science* 103:199-207). Mature sunflower seed (*Helianthus annuus* L.) are dehulled using a single wheat-head thresher. Seeds are surface sterilized for 30 minutes in a 20% Clorox bleach solution with the addition of two drops of Tween 20 per 50 ml of solution. The seeds are rinsed twice with sterile distilled water.

Split embryonic axis explants are prepared by a modification of procedures described by Schrammeijer *et al.* (Schrammeijer *et al.* (1990) *Plant Cell Rep.* 9: 55-60). Seeds are imbibed in distilled water for 60 minutes following the surface sterilization procedure. The cotyledons of each seed are then broken off, producing a clean fracture at the plane of the embryonic axis. Following excision of the root tip, the explants are bisected longitudinally between the primordial leaves. The two halves are placed, cut surface up, on GBA medium consisting of Murashige and Skoog mineral elements (Murashige *et al.* (1962) *Physiol. Plant.*, 15: 473-497), Shepard's vitamin additions (Shepard (1980) in *Emergent Techniques for the Genetic Improvement of Crops*



(University of Minnesota Press, St. Paul, Minnesota), 40 mg/l adenine sulfate, 30 g/l sucrose, 0.5 mg/l 6-benzyl-aminopurine (BAP), 0.25 mg/l indole-3-acetic acid (IAA), 0.1 mg/l gibberellic acid (GA<sub>3</sub>), pH 5.6, and 8 g/l Phytagar.

The explants are subjected to microprojectile bombardment prior to

- 5 *Agrobacterium* treatment (Bidney *et al.* (1992) *Plant Mol. Biol.* 18: 301-313). Thirty to forty explants are placed in a circle at the center of a 60 X 20 mm plate for this treatment. Approximately 4.7 mg of 1.8 mm tungsten microprojectiles are resuspended in 25 ml of sterile TE buffer (10 mM Tris HCl, 1 mM EDTA, pH 8.0) and 1.5 ml aliquots are used per bombardment. Each plate is bombarded twice through a 150 mm nytex screen placed
- 10 2 cm above the samples in a PDS 1000<sup>®</sup> particle acceleration device.

- Disarmed *Agrobacterium tumefaciens* strain EHA105 is used in all transformation experiments. A binary plasmid vector comprising the expression cassette that contains the PR1-C10 gene operably linked to the ubiquitin promoter is introduced into *Agrobacterium* strain EHA105 via freeze-thawing as described by Holsters *et al.* (1978)
- 15 *Mol. Gen. Genet.* 163:181-187. This plasmid further comprises a kanamycin selectable marker gene (i.e, *nptII*). Bacteria for plant transformation experiments are grown overnight (28°C and 100 RPM continuous agitation) in liquid YEP medium (10 gm/l yeast extract, 10 gm/l Bactopeptone, and 5 gm/l NaCl, pH 7.0) with the appropriate antibiotics required for bacterial strain and binary plasmid maintenance. The suspension
- 20 is used when it reaches an OD<sub>600</sub> of about 0.4 to 0.8. The *Agrobacterium* cells are pelleted and resuspended at a final OD<sub>600</sub> of 0.5 in an inoculation medium comprised of 12.5 mM MES pH 5.7, 1 gm/l NH<sub>4</sub>Cl, and 0.3 gm/l MgSO<sub>4</sub>.

- Freshly bombarded explants are placed in an *Agrobacterium* suspension, mixed, and left undisturbed for 30 minutes. The explants are then transferred to GBA medium
- 25 and co-cultivated, cut surface down, at 26°C and 18-hour days. After three days of co-cultivation, the explants are transferred to 374B (GBA medium lacking growth regulators and a reduced sucrose level of 1%) supplemented with 250 mg/l cefotaxime and 50 mg/l kanamycin sulfate. The explants are cultured for two to five weeks on selection and then transferred to fresh 374B medium lacking kanamycin for one to two weeks of continued
- RTA01/2093378v1

development. Explants with differentiating, antibiotic-resistant areas of growth that have not produced shoots suitable for excision are transferred to GBA medium containing 250 mg/l cefotaxime for a second 3-day phytohormone treatment. Leaf samples from green, kanamycin-resistant shoots are assayed for the presence of NPTII by ELISA and for the presence of transgene expression by assaying for PR1-C10 activity.

NPTII-positive shoots are grafted to Pioneer<sup>®</sup> hybrid 6440 *in vitro*-grown sunflower seedling rootstock. Surface sterilized seeds are germinated in 48-0 medium (half-strength Murashige and Skoog salts, 0.5% sucrose, 0.3% gelrite, pH 5.6) and grown under conditions described for explant culture. The upper portion of the seedling is removed, a 1 cm vertical slice is made in the hypocotyl, and the transformed shoot inserted into the cut. The entire area is wrapped with parafilm to secure the shoot. Grafted plants can be transferred to soil following one week of *in vitro* culture. Grafts in soil are maintained under high humidity conditions followed by a slow acclimatization to the greenhouse environment. Transformed sectors of T<sub>0</sub> plants (parental generation) maturing in the greenhouse are identified by NPTII ELISA and/or by PR1-C10 activity analysis of leaf extracts while transgenic seeds harvested from NPTII-positive T<sub>0</sub> plants are identified by PR1-C10 activity analysis of small portions of dry seed cotyledon.

An alternative sunflower transformation protocol allows the recovery of transgenic progeny without the use of chemical selection pressure. Seeds are dehulled and surface-sterilized for 20 minutes in a 20% Clorox bleach solution with the addition of two to three drops of Tween 20 per 100 ml of solution, then rinsed three times with distilled water. Sterilized seeds are imbibed in the dark at 26°C for 20 hours on filter paper moistened with water. The cotyledons and root radical are removed, and the meristem explants are cultured on 374E (GBA medium consisting of MS salts, Shepard vitamins, 40 mg/l adenine sulfate, 3% sucrose, 0.5 mg/l 6-BAP, 0.25 mg/l IAA, 0.1 mg/l GA, and 0.8% Phytagar at pH 5.6) for 24 hours under the dark. The primary leaves are removed to expose the apical meristem, around 40 explants are placed with the apical dome facing upward in a 2 cm circle in the center of 374M (GBA medium with 1.2% Phytagar), and then cultured on the medium for 24 hours in the dark.

RTA01/2093378v1

Approximately 18.8 mg of 1.8 µm tungsten particles are resuspended in 150 µl absolute ethanol. After sonication, 8 µl of it is dropped on the center of the surface of macrocarrier. Each plate is bombarded twice with 650 psi rupture discs in the first shelf at 26 mm of Hg helium gun vacuum.

5 The plasmid of interest is introduced into *Agrobacterium tumefaciens* strain EHA105 via freeze thawing as described previously. The pellet of overnight-grown bacteria at 28 °C in a liquid YEP medium (10 g/l yeast extract, 10 g/l Bactopeptone, and 5 g/l NaCl, pH 7.0) in the presence of 50 µg/l kanamycin is resuspended in an inoculation medium (12.5 mM 2-mM 2-(N-morpholino) ethanesulfonic acid, MES, 1 g/l 10 NH<sub>4</sub>Cl and 0.3 g/l MgSO<sub>4</sub> at pH 5.7) to reach a final concentration of 4.0 at OD 600. Particle-bombarded explants are transferred to GBA medium (374E), and a droplet of bacteria suspension is placed directly onto the top of the meristem. The explants are co-cultivated on the medium for 4 days, after which the explants are transferred to 374C medium (GBA with 1% sucrose and no BAP, IAA, GA3 and supplemented with 250 15 µg/ml cefotaxime). The plantlets are cultured on the medium for about two weeks under 16-hour day and 26 °C incubation conditions.

Explants (around 2 cm long) from two weeks of culture in 374C medium are screened for PR1-C10 activity. After positive (i.e., for PR1-C10 expression) explants are identified, those shoots that fail to exhibit PR1-C10 activity are discarded, and every 20 positive explant is subdivided into nodal explants. One nodal explant contains at least one potential node. The nodal segments are cultured on GBA medium for three to four days to promote the formation of auxiliary buds from each node. Then they are transferred to 374C medium and allowed to develop for an additional four weeks. Developing buds are separated and cultured for an additional four weeks on 374C 25 medium. Pooled leaf samples from each newly recovered shoot are screened again by the appropriate protein activity assay. At this time, the positive shoots recovered from a single node will generally have been enriched in the transgenic sector detected in the initial assay prior to nodal culture.

Recovered shoots positive for PR1-C10 expression are grafted to Pioneer hybrid 6440 *in vitro*-grown sunflower seedling rootstock. The rootstocks are prepared in the following manner. Seeds are dehulled and surface-sterilized for 20 minutes in a 20% Clorox bleach solution with the addition of two to three drops of Tween 20 per 100 ml of solution, and are rinsed three times with distilled water. The sterilized seeds are germinated on the filter moistened with water for three days, then they are transferred into 48 medium (half-strength MS salt, 0.5% sucrose, 0.3% gelrite pH 5.0) and grown at 26 °C under the dark for three days, then incubated at 16-hour-day culture conditions. The upper portion of selected seedling is removed, a vertical slice is made in each hypocotyl, and a transformed shoot is inserted into a V-cut. The cut area is wrapped with parafilm. After one week of culture on the medium, grafted plants are transferred to soil. In the first two weeks, they are maintained under high humidity conditions to acclimatize to a greenhouse environment.

The above examples are provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, patents, and patent applications cited herein are indicative of the level of those skilled in the art to which this invention pertains. All publications, patents, and patent applications are hereby incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.